



**CAHIER 08-2001**

**SIMULATION-BASED FINITE-SAMPLE TESTS FOR  
HETEROSKEDASTICITY AND ARCH EFFECTS**

Jean-Marie DUFOUR, Lynda KHALAF,  
Jean-Thomas BERNARD and Ian GENEST

**Centre de recherche  
et développement en économie**

C.P. 6128, succursale Centre-ville  
Montréal QC H3C 3J7

Téléphone : (514) 343-6557  
Télécopieur : (514) 343-5831  
[crde@crde.umontreal.ca](mailto:crde@crde.umontreal.ca)  
<http://www.crde.umontreal.ca/>

Université   
de Montréal

CAHIER 08-2001

**SIMULATION-BASED FINITE-SAMPLE TESTS FOR  
HETEROSKEDASTICITY AND ARCH EFFECTS**

Jean-Marie DUFOUR<sup>1</sup>, Lynda KHALAF<sup>2</sup>,  
Jean-Thomas BERNARD<sup>3</sup>, and Ian GENEST<sup>3</sup>

- <sup>1</sup> Centre de recherche et développement en économie (C.R.D.E.) and Département de sciences économiques, Université de Montréal, and CIRANO
- <sup>2</sup> C.R.D.E. and Département de sciences économiques, Université de Montréal, and GREEN, Université Laval
- <sup>3</sup> GREEN, Université Laval

April 2001

---

The authors thank Marie-Claude Beaulieu, Bryan Campbell, Judith Giles and Victoria Zinde-Walsh for several useful comments, and Jean-François Bilodeau for research assistance. This work was supported by the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec). This paper was also partly written at the Centre de Recherche en Économie et Statistique (INSEE, Paris) and the Technische Universität Dresden (Fakultät Wirtschaftswissenschaften).

## RÉSUMÉ

Un grand éventail de tests d'hétéroscédasticité a été proposé en économétrie et en statistique. Bien qu'il existe quelques tests d'homoscédasticité exacts, les procédures couramment utilisées sont généralement fondées sur des approximations asymptotiques qui ne procurent pas un bon contrôle du niveau dans les échantillons finis. Plusieurs études récentes ont tenté d'améliorer la fiabilité des tests d'hétéroscédasticité usuels, sur base de méthodes de type Edgeworth, Bartlett, jackknife et bootstrap. Cependant, ces méthodes demeurent approximatives. Dans cet article, nous décrivons une solution au problème de contrôle du niveau des tests d'homoscédasticité dans les modèles de régression linéaire. Nous étudions des procédures basées sur les critères de test standard [e.g., les critères de Goldfeld-Quandt, Glejser, Bartlett, Cochran, Hartley, Breusch-Pagan-Godfrey, White et Szroeter], de même que des tests pour l'hétéroscédasticité autorégressive conditionnelle (les modèles de type ARCH). Nous suggérons plusieurs extensions des procédures usuelles (les statistiques de type-sup ou combinées) pour tenir compte de points de ruptures inconnus dans la variance des erreurs. Nous appliquons la technique des tests de Monte Carlo (MC) de façon à obtenir des seuils de signification marginaux (les valeurs- $p$ ) exacts pour les tests usuels et les nouveaux tests que nous proposons. Nous démontrons que la procédure de MC permet de résoudre les problèmes des distributions compliquées sous l'hypothèse nulle, en particulier ceux associés aux statistiques de type-sup, aux statistiques combinées et aux paramètres de nuisance non-identifiés sous l'hypothèse nulle. La méthode proposée fonctionne exactement de la même manière en présence de lois Gaussiennes et non Gaussiennes [comme par exemple les lois aux queues épaisses ou les lois stables]. Nous évaluons la performance des procédures proposées par simulation. Les expériences de Monte Carlo que nous effectuons portent sur : (1) les alternatives de type ARCH, GARCH et ARCH-en-moyenne; (2) le cas où la variance augmente de manière monotone en fonction : (i) d'une variable exogène, et (ii) de la moyenne de la variable dépendante; (3) l'hétéroscédasticité groupée; (4) les ruptures en variance à des points inconnus. Nos résultats montrent que les tests proposés permettent de contrôler parfaitement le niveau et ont une bonne puissance.

Mots clés : hétéroscédasticité, homoscédasticité, régression linéaire, test de Monte Carlo, test exact, test valide en échantillon fini, test de spécification, ARCH, GARCH, ARCH-en-moyenne, distribution stable, stabilité structurelle

## ABSTRACT

A wide range of tests for heteroskedasticity have been proposed in the econometric and statistics literature. Although a few exact homoskedasticity tests are available, the commonly employed procedures are quite generally based on asymptotic approximations which may not provide good size control in finite samples. There has been a number of recent studies that seek to improve the reliability of common heteroskedasticity tests using Edgeworth, Bartlett, jackknife and bootstrap methods. Yet the latter remain approximate. In this paper, we describe a solution to the problem of controlling the size of homoskedasticity tests in linear regression contexts. We study procedures based on the standard test statistics [e.g., the Goldfeld-Quandt, Glejser, Bartlett, Cochran, Hartley, Breusch-Pagan-Godfrey, White and Szroeter criteria] as well as tests for autoregressive conditional heteroskedasticity (ARCH-type models). We also suggest several extensions of the existing procedures (sup-type of combined test statistics) to allow for unknown breakpoints in the error variance. We exploit the technique of Monte Carlo tests to obtain provably exact  $p$ -values, for both the standard and the new tests suggested. We show that the MC test procedure conveniently solves the intractable null distribution problem, in particular those raised by the sup-type and combined test statistics as well as (when relevant) unidentified nuisance parameter problems under the null hypothesis. The method proposed works in exactly the same way with both Gaussian and non-Gaussian disturbance distributions [such as heavy-tailed or stable distributions]. The performance of the procedures is examined by simulation. The Monte Carlo experiments conducted focus on : (1) ARCH, GARCH, and ARCH-in-mean alternatives; (2) the case where the variance increases monotonically with : (i) one exogenous variable, and (ii) the mean of the dependent variable; (3) grouped heteroskedasticity; (4) breaks in variance at unknown points. We find that the proposed tests achieve perfect size control and have good power.

Key words : heteroskedasticity, homoskedasticity, linear regression, Monte Carlo test, exact test, finite-sample test, specification test, ARCH, GARCH, ARCH-in-mean, stable distribution, structural stability

# Contents

<b>List of Propositions</b>	<b>iv</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Framework</b>	<b>4</b>
<b>3. Test statistics</b>	<b>5</b>
3.1. Tests based on auxiliary regressions . . . . .	6
3.1.1. Standard auxiliary regression tests . . . . .	6
3.1.2. Auxiliary regression tests against an unknown variance breakpoint . . . . .	7
3.2. Tests against ARCH-type heteroskedasticity . . . . .	8
3.3. Tests based on grouping . . . . .	10
3.3.1. Goldfeld-Quandt tests against an unknown variance breakpoint . . . . .	10
3.3.2. Generalized Bartlett tests . . . . .	12
3.3.3. Szroeter-type tests . . . . .	12
3.3.4. Generalized Cochran-Hartley tests . . . . .	14
3.3.5. Grouping tests against a mean dependent variance . . . . .	14
<b>4. Finite-sample distributional theory</b>	<b>15</b>
<b>5. Simulation experiments</b>	<b>18</b>
5.1. Tests for ARCH and GARCH effects . . . . .	18
5.2. Tests of variance as a linear function of exogenous variables . . . . .	24
5.2.1. Level . . . . .	27
5.2.2. Power . . . . .	27
5.3. Grouped heteroskedasticity . . . . .	27
5.3.1. Level . . . . .	28
5.3.2. Power . . . . .	28
5.4. Tests for break in variance . . . . .	32
<b>6. Conclusion</b>	<b>36</b>

## List of Propositions

4.1	<b>Proposition :</b> Characterization of pivotal statistics . . . . .	15
4.2	<b>Proposition :</b> Pivotal property of residual-based statistics . . . . .	16

## List of Tables

1	Survey of empirical literature on the use heteroskedasticity tests . . . . .	2
2	Parameter values used for the GARCH models . . . . .	19
3	Testing for ARCH and GARCH . . . . .	19
4	Empirical size of ARCH-M tests . . . . .	20
5	Power of MC ARCH-M tests: normal errors and D1 design . . . . .	21
6	Power of MC ARCH-M tests: various error distributions and D2 design . . . . .	22
7	Variance proportional to a regressor . . . . .	25
8	Variance as a function of the mean . . . . .	26
9	Grouped heteroskedasticity . . . . .	29
10	Break in variance at unknown points . . . . .	33

## 1. Introduction

Detecting and making adjustments for the presence of heteroskedasticity in the disturbances of statistical models is one of the fundamental problems of econometric methodology. We study here the problem of testing the homoskedasticity of linear regression disturbances, under parametric (possibly non-Gaussian) distributional assumptions, against a wide range of alternatives, especially in view of obtaining more reliable or more powerful procedures. The heteroskedastic schemes we consider include random volatility models, such as ARCH and GARCH error structures, variances which are functions of exogenous variables, as well as discrete breaks at (possibly unknown) points.

The statistical and econometric literatures on testing for heteroskedasticity is quite extensive; for reviews, the reader may consult Judge, Griffiths, Carter Hill, Lütkepohl, and Lee (1985), Godfrey (1988), Pagan and Pak (1993) and Davidson and MacKinnon (1993, Chapters 11 and 16). In linear regression contexts, the most popular procedures include the Goldfeld-Quandt  $F$ -test [Goldfeld and Quandt (1965)], Glejser's regression-type tests [Glejser (1969)], Ramsey's versions of the Bartlett (1937) test [Ramsey (1969)], the Breusch-Pagan-Godfrey Lagrange multiplier (LM) test [Godfrey (1978), Breusch and Pagan (1979)], White's general test [White (1980)], Koenker's studentized test [Koenker (1981)], and Cochran-Hartley-type tests against grouped heteroskedasticity [Cochran (1941), Hartley (1950), Rivest (1986)]; see the literature survey results in Table 1. Other proposed methods include likelihood (LR) tests against specific alternatives [see, for example, Harvey (1976), Buse (1984), Maekawa (1988) or Binkley (1992)] and "robust procedures", such as the Goldfeld and Quandt (1965) peak test and the procedures suggested by Bickel (1978), Koenker and Bassett (1982) and Newey and Powell (1987).

The above methods do not usually take variances as a primary object of interest, but as nuisance parameters that must be taken into account (and eventually eliminated) when making inference on other model parameters (such as regression coefficients). More recently, in time series contexts and especially financial data analysis, the modeling of variances (volatilities) as a stochastic process has come to be viewed also as an important aspect of data analysis, leading to the current popularity of ARCH, GARCH and other similar models; see Engle (1982, 1995), Engle, Hendry, and Trumble (1985), Bollerslev, Engle, and Nelson (1994), LeRoy (1996), Palm (1996), and Gouriéroux (1997). As a result, detecting the presence of conditional stochastic heteroskedasticity has become an important issue, and a number of tests against the presence of such effects have been proposed; see Engle (1982), Lee and King (1993), Bera and Ra (1995) and Hong and Shehadeh (1999).

Despite the large spectrum of tests available, the vast majority of the proposed procedures are based on large-sample approximations, even when it is assumed that the disturbances are independent and identically distributed (i.i.d.) with a normal distribution under the null hypothesis. So there has been a number of recent studies that seek to improve the finite-sample reliability of commonly used homoskedasticity tests. In particular, Honda (1988) and Cribari-Neto and Ferrari (1995) derived Edgeworth and Bartlett modifications for the Breusch-Pagan-Godfrey criteria, while Cribari-Neto and Zarkos (1998) considered bootstrap versions of the latter procedures. Tests based on the jackknife method have also been considered; see, for example, Giaccotto and Sharma (1988) and Sharma and Giaccotto (1991). In a multi-equations framework, Bewley and Theil (1987) suggested a simulation-based test for a particular testing problem; however, they did not supply a distributional

Table 1. Survey of empirical literature on the use heteroskedasticity tests

Heteroskedasticity test used	Literature Share
Tests for ARCH and GARCH effects	25.3%
Breusch-Pagan-Godfrey-Koenker	20.9%
White's test	11.3%
Goldfeld-Quandt	6.6%
Glejser's test	2.9%
Hartley's test	0.3%
Other tests	1.9%
Use of heteroskedasticity consistent standard errors	30.3%

Note \_ This survey is based on 379 papers published in The Journal of Business and Economic Statistics, The Journal of Applied Econometrics, Applied Economics, the Canadian Journal of Economics, Economics Letters, over the period 1980 -1997. These results were generously provided by Judith Giles.

theory, either exact or asymptotic.

A limited number of provably exact heteroskedasticity tests, for which the level can be controlled for any given sample size, have been suggested. These include: (1) the familiar Goldfeld-Quandt  $F$ -test and its extensions based on BLUS [Theil (1971)] and recursive residuals [Harvey and Phillips (1974)], which are built against a very specific (two-regime) alternative; (2) a number of procedures in the class introduced by Szroeter (1978), which also include Goldfeld-Quandt-type tests as a special case [see Harrison and McCabe (1979), Harrison (1980, 1981, 1982), King (1981) and Evans and King (1985a)]; (3) the procedures proposed by Evans and King (1985b) and McCabe (1986). All these tests are specifically designed to apply under the assumption that regression disturbances are independent and identically distributed (i.i.d.) according to a normal distribution under the null hypothesis. Further, except for the Goldfeld-Quandt procedure, these tests require techniques for computing the distributions of general quadratic forms in normal variables such as the Imhof (1961) method, and they are seldom used (see Table 1).

Several studies compare various heteroskedasticity tests from the reliability and power viewpoints; see, for example, Ali and Giaccotto (1984), Buse (1984), MacKinnon and White (1985), Griffiths and Surekha (1986), Farebrother (1987), Evans (1992), Godfrey (1996), and, in connection with GARCH tests, Engle, Hendry, and Trumble (1985), Lee and King (1993), Sullivan and Giles (1995), Bera and Ra (1995) and Lumsdaine (1995). In addition, most of the references cited above include Monte Carlo evidence on the relative performance of various tests. The main findings that emerge from these studies are the following: (i) no single test has the greatest power against all alternatives; (ii) tests based on OLS residuals perform best; (iii) the actual level of asymptotically justified tests is often quite far from the nominal level: some are over-sized [see, for example, Honda (1988), Ali and Giaccotto (1984) and Binkley (1992)], while others are heavily under-sized, leading to important power losses [see Lee and King (1993), Evans (1992), Honda (1988), Griffiths and Surekha (1986), and Binkley (1992)]; (iv) the incidence of inconclusiveness is high among the bounds tests; (v) the exact tests compare favorably with asymptotic tests but can be quite difficult to



implement in practice. Of course, these conclusions may be influenced by the special assumptions and simulation designs that were considered.

In this paper, we describe a general solution to the problem of controlling the size of homoskedasticity tests in linear regression contexts. We consider procedures based on the standard tests as well as several extensions of the latter. Specifically, we focus on the following heteroskedastic alternatives: (1) ARCH, GARCH and ARCH-in-mean (ARCH-M) effects; (2) breaks in variance at possibly unknown points; (3) cases where the variance depends on a vector of exogenous variables; (4) the case where the variance is a function of the mean of the dependent variable; (5) grouped heteroskedasticity. We exploit the technique of Monte Carlo (MC) tests [Dwass (1957), Barnard (1963), Jöckel (1986), Dufour and Kiviet (1996, 1998)] to obtain provably exact randomized analogues of the tests considered. This simulation-based procedure yields an exact test whenever the distribution of the test statistic does not depend on unknown nuisance parameters (i.e., it is *pivotal*) under the null hypothesis. The fact that the relevant analytical distributions are quite complicated is not a problem in this context: all we need is the possibility of simulating the relevant test statistic under the null hypothesis. In particular, this covers many cases where the finite-sample distribution of the test statistic is intractable or involves parameters which are unidentified under the null hypothesis, as occurs in the problems studied by Davies (1977, 1987), Andrews and Ploberger (1995), and Hansen (1996). Further the method allows one to consider any error distribution that can be simulated, which of course covers both Gaussian and many non-Gaussian distributions (such as stable distributions).

We show here that all the standard homoskedasticity test statistics considered are indeed pivotal. In particular, we observe that a large class of residual-based tests for heteroskedasticity [studied from an asymptotic viewpoint by Pagan and Hall (1983)] are pivotal in finite samples, hence allowing the construction of finite-sample MC versions of these. In this way, the size of many popular asymptotic procedures, such as the Breusch-Pagan-Godfrey, White, Glejser, Bartlett, and Cochran-Hartley-type tests, can be perfectly controlled for any parametric error distribution (Gaussian or non-Gaussian) specified up to an unknown scale parameter. Tests for which a finite-sample theory has been supplied for Gaussian distributions, such as the Goldfeld-Quandt and various Szroeter-type tests, are extended to allow for non-Gaussian distributions. Further, we show that various bounds procedures that were proposed to deal with intractable finite-sample distributions [e.g., by Szroeter (1978), King (1981) and McCabe (1986)] can be avoided altogether in this way.

Our results also cover the important problem of testing for ARCH, GARCH and ARCH-M effects. In this case, MC tests provide finite-sample homoskedasticity tests against standard ARCH-type alternatives where the noise that drives the ARCH process is i.i.d. Gaussian, and allow one to deal in a similar way with non-Gaussian disturbances. In non-standard test problems, such as the ARCH-M case, we observe that the MC procedure circumvents the unidentified nuisance parameter problem. Further, due to the convenience of MC test methods, we define a number of new test statistics and show how they can be implemented. These include: (1) combined Breusch-Pagan-Godfrey tests against a break in the variance at an unknown date (or point); (2) combined Goldfeld-Quandt tests against a variance break at an unspecified point, based on the minimum (*sup*-type) or the product of individual *p*-values; (3) extensions of the classic Cochran (1941) and Hartley (1950) tests, against grouped heteroskedasticity, to the regression framework using pooled regression residuals.

Although the null distributions of many of these tests may be quite difficult to establish in finite samples and even asymptotically, we show that the tests can easily be implemented as finite-sample MC tests.<sup>1</sup>

To assess the validity of residual-based homoskedasticity tests, Godfrey (1996, section 2) defined the notion of “robustness to estimation effects”. In principle, a test is considered robust to estimation effects if the underlying asymptotic distribution is the same irrespective of whether disturbances or residuals are used to construct the test statistic. Our approach to residual-based tests departs from Godfrey’s asymptotic framework. As noted earlier, MC homoskedasticity tests are based on a finite-sample distributional theory. Indeed, since the test criteria considered are pivotal under the null hypothesis, we shall be able to control perfectly type I error probabilities whenever the error distribution is specified up to an unknown scale parameter [e.g., the variance], even with non-normal errors. Therefore, the adjustments proposed by Godfrey (1996) or Koenker (1981) are not necessary for controlling size. It is also of interest to note that Breusch and Pagan (1979) recognized the pivotal property of the LM homoskedasticity test which might be exploited to obtain simulation-based cut-off points. Here we provide a clear simulation-based strategy that allows one to control the size of the tests even with a very small number of replications.

The performance of the proposed tests is examined by simulation. The MC studies we consider assess the various tests, assuming a correctly specified model. We do not address the effects on the tests which result from misspecifying the model and/or the testing problem. Our results indicate that the MC versions of the popular tests typically have superior size and power properties.

The paper is organized as follows. Section 2 presents the statistical framework and Section 3 defines the test criteria considered. The Monte Carlo test procedure is described in Section 4. In Section 5, we report the results of the Monte Carlo experiments. Section 6 concludes.

## 2. Framework

We consider the linear model

$$y_t = x_t' \beta + u_t, \quad (2.1)$$

$$u_t = \sigma_t \varepsilon_t, \quad t = 1, \dots, T, \quad (2.2)$$

where  $x_t = (x_{t1}, x_{t2}, \dots, x_{tk})'$ ,  $X \equiv [x_1, \dots, x_T]'$  is a full-column rank  $T \times k$  matrix,  $\beta = (\beta_1, \dots, \beta_k)'$  is a  $k \times 1$  vector of unknown coefficients,  $\sigma_1, \dots, \sigma_T$  are (possibly random) scale parameters, and

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)' \text{ is a random vector with a completely specified continuous distribution conditional on } X. \quad (2.3)$$

---

<sup>1</sup>For example, the combined test procedures proposed here provide solutions to a number of *change-point* problems. For further discussion of the related distributional issues, the reader may consult MacNeill (1978), Shaban (1980), Chu and White (1992), Zivot and Andrews (1992), Andrews (1993) and Hansen (1997)]

Clearly the case where the disturbances are normally distributed is included as a special case. We are concerned with the problem of testing the null hypothesis

$$H_0 : \sigma_t^2 = \sigma^2, t = 1, \dots, T, \text{ for some } \sigma, \quad (2.4)$$

against the alternative  $H_A : \sigma_t^2 \neq \sigma_s^2$ , for at least one value of  $t$  and  $s$ .

The hypothesis defined by (2.1) - (2.4) does not preclude dependence nor heterogeneity among the components of  $\varepsilon$ . So in most cases of practical interest, one would further restrict the distribution of  $\varepsilon$ , for example by assuming that the elements of  $\varepsilon$  are independent and identically distributed (i.i.d.), i.e.

$$\varepsilon_1, \dots, \varepsilon_T \text{ are i.i.d. according to some given distribution } F_0, \quad (2.5)$$

which entails that  $u_1, \dots, u_T$  are i.i.d. with distribution function  $P[u_t \leq v] = F_0(v/\sigma)$  under  $H_0$ . In particular, it is quite common to assume that

$$\varepsilon_1, \dots, \varepsilon_T \stackrel{i.i.d.}{\sim} N[0, 1], \quad (2.6)$$

which entails that  $u_1, \dots, u_T$  are i.i.d.  $N[0, \sigma^2]$  under  $H_0$ . However, as shown in Section 4, the normality assumption is not needed for several of our results; in particular, it is not at all required for the validity of MC tests for general hypotheses of the form (2.1) - (2.4), hence, a fortiori, if (2.4) is replaced by the stronger assumption (2.5) or (2.6).

We shall focus on the following special cases of heteroskedasticity ( $H_A$ ), namely:

$H_1$  : GARCH and ARCH-M alternatives;

$H_2$  :  $\sigma_t^2$  depends monotonically on a linear function  $z_t' \alpha$  of a vector  $z_t$  of exogenous variables;

$H_3$  :  $\sigma_t^2$  is a monotonic function of  $E(y_t)$  (or  $|E(y_t)|$ );

$H_4$  :  $\sigma_t^2$  is the same within  $p$  subsets of the data but differs across the subsets; the latter specification is frequently termed *grouped heteroskedasticity*. Note that  $H_4$  may include the hypothesis that the variance changes discretely at some point in time (which may be specified or not).

In most cases, the tests considered are functions of the least squares residuals

$$\hat{u} = (\hat{u}_1, \dots, \hat{u}_T)' = y - X\hat{\beta} \quad (2.7)$$

where  $\hat{\beta} = (X'X)^{-1}X'y$  denotes the ordinary least squares (OLS) estimate of  $\beta$ . We shall also write:

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2. \quad (2.8)$$

### 3. Test statistics

As already mentioned, among the numerous tests for heteroskedasticity which have been proposed, they are nowadays quite unevenly used [see Table 1], so we have tried to concentrate on the most

popular procedures and alternatives. Unless stated otherwise, we shall assume in this section that (2.6) holds, even though the asymptotic distributional theory for several of the proposed procedures can be obtained under weaker assumptions. The tests we shall study can be conveniently classified in three (not mutually exclusive) categories: (i) the general class of tests based on an auxiliary regression involving OLS residuals and some vector of explanatory variables  $z_t$  for the error variance; (ii) tests against ARCH-type alternatives; (iii) tests against grouped heteroskedasticity.

### 3.1. Tests based on auxiliary regressions

#### 3.1.1. Standard auxiliary regression tests

To introduce these tests in their simplest form, consider the following auxiliary regressions:

$$\hat{u}_t^2 = z_t' \alpha + w_t, \quad t = 1, \dots, T, \quad (3.1)$$

$$\hat{u}_t^2 - \hat{\sigma}^2 = z_t' \alpha + w_t, \quad t = 1, \dots, T, \quad (3.2)$$

$$|\hat{u}_t| = z_t' \alpha + w_t, \quad t = 1, \dots, T, \quad (3.3)$$

where  $z_t = (1, z_{t2}, \dots, z_{tm})'$  is a vector of  $m$  fixed regressors on which  $\sigma_t$  may depend,  $\alpha = (\alpha_1, \dots, \alpha_m)'$  and  $w_t$ ,  $t = 1, \dots, T$ , are treated as error terms.<sup>2</sup> The Breusch-Pagan-Godfrey (*BPG*) LM criterion [Breusch and Pagan (1979), Godfrey (1978)] may be obtained as the explained sum of squares (*ESS*) from the regression associated with (3.1) divided by  $2\hat{\sigma}^4$ . The Koenker (*K*) test statistic [Koenker (1981)] is  $T$  times the uncentered  $R^2$  from regression (3.2). White's (*W*) test statistic is  $T$  times the uncentered  $R^2$  from regression (3.1) using for  $z_t$  the  $r \times 1$  observations on the non redundant variables in the vector  $x_t \otimes x_t$ . These tests can be derived as LM-type tests against alternatives of the form

$$H_A : \sigma_t^2 = g(z_t' \alpha) \quad (3.4)$$

where  $g(\cdot)$  is a twice differentiable function. Under  $H_0$  and standard asymptotic regularity conditions, we have:

$$BPG \overset{asy}{\rightsquigarrow} \chi^2(m-1),$$

$$K \overset{asy}{\rightsquigarrow} \chi^2(m-1),$$

$$W \overset{asy}{\rightsquigarrow} \chi^2(r-1),$$

where the symbol  $\overset{asy}{\rightsquigarrow}$  indicates that the test statistic is asymptotically distributed as indicated (under  $H_0$  as  $T \rightarrow \infty$ ). The standard  $F$  statistic to test  $\alpha_2 = \dots = \alpha_m = 0$  in the context of (3.3) yields the Glejser (*G*) test [Glejser (1969)]. Again, under  $H_0$  and standard regularity conditions,

$$(T-k)G \overset{asy}{\rightsquigarrow} \chi^2(m-1).$$

Below, we shall also consider  $F(m-1, T-k)$  distribution as an approximation to the null distribution of this statistic. Honda (1988) has also provided a size-correction formula for the *BPG*

---

<sup>2</sup>See Pagan and Hall (1983) for a formal regression interpretation of the tests.

statistic. White's test was designed against the general alternative  $H_A$ . The above version of the Glejser test is valid for the special case where the variance is proportional to  $z_t'\alpha$ .

Godfrey (1996) has recently shown that, unless the error distribution is symmetric, the  $G$  test is deficient in the following sense. The residual-based test is not asymptotically equivalent to a conformable  $\chi^2$  test based on the true errors. Therefore, the  $G$  test may not achieve size control. We will show below that this difficulty is circumvented by our proposed MC version of the test. In the same vein, we argue that from a MC test perspective, choosing the Koenker statistic rather than the  $BPG$  has no incidence on size control. We provide a rigorous justification for the latter arguments in Section 4.

Tests against discrete breaks in variance at some specified date  $\tau$  may be applied in the above framework by defining  $z_t$  as a dummy variable of the form  $z_t = z_t(\tau)$ , where

$$z_t(\tau) = \begin{cases} 0, & t \leq \tau \\ 1, & t > \tau \end{cases} . \quad (3.5)$$

Pagan and Hall (1983, p. 117) provide the relevant special form of the  $BPG$  test statistic. We also suggest here extensions of this procedure to the case where the break-date  $\tau$  is left unspecified, and thus may take any one of the values  $\tau = 1, \dots, T - 1$ . One may then compute a different test statistic for each one of these possible break-dates. Note the problem of combining inference based on the resulting multiple tests was not solved by Pagan and Hall (1983).

### 3.1.2. Auxiliary regression tests against an unknown variance breakpoint

Let  $BPG_\tau$  be the  $BPG$  statistic obtained on using  $z_t = z_t(\tau)$ , where  $\tau = 1, \dots, T - 1$ . When used as a single test, the  $BPG_\tau$  statistic is significant at level  $\alpha$  when  $BPG_\tau \geq \chi_\alpha^2(1)$ , or equivalently when  $G_{\chi_1}(BPG_\tau) \leq \alpha$ , where  $\chi_\alpha^2(1)$  solves the equation  $G_{\chi_1}[\chi_\alpha^2(1)] = \alpha$  and  $G_{\chi_1}(x) = P[\chi^2(1) \geq x]$  is the survival function of the  $\chi^2(1)$  probability distribution.  $G_{\chi_1}(BPG_\tau)$  is the asymptotic  $p$ -value associated with  $BPG_\tau$ . We propose here two methods for combining the  $BPG_\tau$  tests.

The first one rejects  $H_0$  when at least one of the  $p$ -values for  $\tau \in J$  is sufficiently small, where  $J$  is some appropriate subset of the time interval  $\{1, 2, \dots, T - 1\}$ , such as  $J = [\tau_1, \tau_2]$  where  $1 \leq \tau_1 < \tau_2 \leq T - 1$ . In theory,  $J$  may be any non-empty subset of  $\{1, 2, \dots, T - 1\}$ . More precisely, we reject  $H_0$  at level  $\alpha$  when  $pv_{\min}(BPG; J) \leq p_0(\alpha; J)$  where

$$pv_{\min}(BPG; J) \equiv \min\{G_{\chi_1}(BPG_\tau) : \tau \in J\} \quad (3.6)$$

and  $p_0(\alpha; J)$  is the largest point such that  $P[pv_{\min}(BPG; J) \leq p_0(\alpha; J)] \leq \alpha$  under  $H_0$ , or equivalently when  $F_{\min}(BPG; J) \geq F_{\min}(\alpha; J)$  where

$$F_{\min}(BPG; J) \equiv 1 - \min\{G_{\chi_1}(BPG_\tau) : \tau \in J\} \quad (3.7)$$

and  $F_{\min}(\alpha; J) = 1 - p_0(\alpha; J)$ . In general, to avoid over-rejecting,  $p_0(\alpha; J)$  should be smaller than  $\alpha$ . This method of combining tests was suggested by Tippett (1931) and Wilkinson (1951) in the case

of independent test statistics. Here, it is however clear that the statistics  $BPG_\tau$ ,  $\tau = 1, \dots, T-1$ , are not independent, with possibly a complex dependence structure.

The second method we consider consists in rejecting  $H_0$  when the product (rather than the minimum) of the  $p$ -values  $pv_\times(BPG; J) \equiv \prod_{\tau \in J} G_{\chi_1}(BPG_\tau)$  is small, or equivalently when  $F_\times(BPG; J) \geq \bar{F}_\times(J; \alpha)$  where

$$F_\times(BPG; J) \equiv 1 - \prod_{\tau \in J} G_{\chi_1}(BPG_\tau) \quad (3.8)$$

and  $\bar{F}_\times(J; \alpha)$  is the largest point such that  $P[F_\times(BPG; J) \geq \bar{F}_\times(J; \alpha)] \leq \alpha$  under  $H_0$ . This general method of combining  $p$ -values was originally suggested by Fisher (1932) and Pearson (1933), again for independent test statistics.<sup>3</sup> We also propose here to consider a modified version of  $F_\times(BPG; J)$  based on a subset of the  $p$ -values  $G_{\chi_1}(BPG_\tau)$ . Specifically, we shall consider a variant of  $F_\times(BPG; J)$  based on the four smallest  $p$ -values:

$$F_\times(BPG; \hat{J}_{(4)}) = 1 - \prod_{\tau \in \hat{J}_{(4)}} G_{\chi_1}(BPG_\tau) \quad (3.9)$$

where  $\hat{J}_{(4)}$  is the set of the four smallest  $p$ -values in the series  $\{G_{\chi_1}(BPG_\tau) : \tau = 1, 2, \dots, T-1\}$ . We shall see later that this modified statistic has better power properties. Implicitly, the maximal number of  $p$ -values retained (four in this case) may be chosen to reflect (prior) knowledge on potential break dates.

We will see below that the technique of MC tests provides a simple way of controlling the size of the tests  $F_{\min}(BPG; J)$ ,  $F_\times(BPG; J)$  and  $F_\times(BPG; \hat{J}_{(4)})$ , although their finite-sample and even their asymptotic distributions may be quite intractable.

### 3.2. Tests against ARCH-type heteroskedasticity

In the context of conditional heteroskedasticity, artificial regressions provide an easy way to compute tests for GARCH effects. Engle (1982) proposed a LM test based on the following framework:

$$\begin{aligned} y_t &= x_t' \beta + u_t, \quad t = 1, \dots, T, \\ u_t |_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2, \end{aligned} \quad (3.10)$$

---

<sup>3</sup>For further discussion of methods for combining tests, the reader may consult Miller (1981), Folks (1984), Savin (1984), Dufour (1989, 1990), Westfall and Young (1993), Gouriéroux and Monfort (1995, Chapter 19), and Dufour and Torrès (1998, 2000).

where  $|_{t-1}$  denotes conditioning of information up to and including  $t - 1$ . The hypothesis of homoskedasticity may then be formulated as follows:

$$H_0 : \alpha_1 = \dots = \alpha_q = 0. \quad (3.11)$$

The Engle test statistic is given by  $TR^2$ , where  $T$  is the sample size,  $R^2$  is the coefficient of determination in the regression of squared OLS residuals  $\hat{u}_t^2$  on a constant and  $\hat{u}_{t-i}^2$  ( $i = 1, \dots, q$ ). Under standard regularity conditions, the asymptotic null distribution of this statistic is  $\chi^2(q)$ . Lee (1991) has also shown that the same test is appropriate against GARCH( $p, q$ ) alternatives, *i.e.*

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \theta_i \sigma_{t-i}^2 + \sum_{i=1}^q \alpha_i u_{t-i}^2, \quad (3.12)$$

and the null hypothesis is

$$H_0 : \alpha_1 = \dots = \alpha_q = \theta_1 = \dots = \theta_p = 0.$$

Lee and King (1993) proposed an alternative (G)ARCH test which exploits the one sided nature of  $H_A$ . The test statistic is

$$LK = \frac{\left\{ (T-q) \sum_{t=q+1}^T [(\hat{u}_t^2 / \hat{\sigma}^2 - 1)] \sum_{i=1}^q \hat{u}_{t-i}^2 \right\} / \left\{ \sum_{t=q+1}^T (\hat{u}_t^2 / \hat{\sigma}^2 - 1)^2 \right\}^{1/2}}{\left\{ (T-q) \sum_{t=q+1}^T \left( \sum_{i=1}^q \hat{u}_{t-i}^2 \right)^2 - \left( \sum_{t=q+1}^T \left( \sum_{i=1}^q \hat{u}_{t-i}^2 \right) \right)^2 \right\}^{1/2}} \quad (3.13)$$

and its asymptotic null distribution is standard normal.<sup>4</sup>

In this paper, we also consider tests against ARCH-M heteroskedasticity (where the shocks affecting the conditional variance of  $y_t$  also have an effect on its conditional mean). This model is an extension of (3.10) which allows the conditional mean to be dependent on the time-varying conditional variance. Formally, the model may be defined as follows:

$$\begin{aligned} y_t &= x_t' \beta + \sigma_t \phi + u_t, \quad t = 1, \dots, T, \\ u_t |_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2. \end{aligned}$$

Then the LM statistic for homoskedasticity hypothesis (3.11) for given  $\phi$  is:

$$LM(\phi) = \frac{1}{2 + \phi^2} \hat{\gamma}' V \left[ V' V - \frac{\phi^2}{2 + \phi^2} V' X (X' X)^{-1} X' V \right]^{-1} V' \hat{\gamma} \quad (3.14)$$

---

<sup>4</sup>We have applied the corrected test formula provided by Lee and King in 1994 [equation (13)].

where  $\hat{\gamma}$  is a  $T \times 1$  vector with elements

$$\hat{\gamma}_t = [(\hat{u}_t/\hat{\sigma})^2 - 1] + \phi\hat{u}_t/\hat{\sigma}$$

and  $V$  is a  $T \times (q + 1)$  matrix whose  $t$ -th row is

$$V_t = (1, \hat{u}_{t-1}^2, \dots, \hat{u}_{t-q}^2);$$

see Bera and Ra (1995). In this case, under  $H_0$ , the parameter  $\phi$  is unidentified. Only the sum of  $\phi$  and the intercept  $(\phi + \beta_1)$  is identifiable under  $H_0$ , although an “estimate” of  $\phi$  can be produced under both the null and the alternative hypotheses. In practice, the latter estimate is substituted for  $\phi$  to implement the LM test using a cut-off point from the  $\chi^2(q)$  distribution. Bera and Ra (1995) also discuss the application of the Davies sup-LM test to this problem and show that this leads to more reliable inference.<sup>5</sup> It is clear, however, that the asymptotic distribution required is quite complicated. We will show below that the MC test procedure may be applied to this sup-LM test. The unidentified nuisance parameter is not a problem for implementing the MC version of the test. Indeed, it is easy to see that the statistic’s finite sample null distribution is nuisance-parameter-free. The simulation experiment in Section 5.1 shows that this method works very well in terms of size and power.

### 3.3. Tests based on grouping

An alternative class of tests assumes that observations can be ordered so that the variance is non-decreasing. In practice, the data are typically sorted according to time or some regressor. In the case of  $H_3$ , the ranking may be based on  $\hat{y}_t$ ; yet this choice may affect the finite-sample null distributions of the test statistics. For further reference, let  $\hat{u}_{(t)}$ ,  $t = 1, \dots, T$ , denote the OLS residuals obtained after reordering the observations (if needed).

#### 3.3.1. Goldfeld-Quandt tests against an unknown variance breakpoint

The most familiar test in this class is the Goldfeld and Quandt (1965, GQ) test which involves separating the ordered sample into three subsets and computing separate OLS regressions on the first and last data subsets. Let  $T_i$ ,  $i = 1, 2, 3$ , denote the number of observations in each of these subsets ( $T = T_1 + T_2 + T_3$ ). Under (2.1) - (2.6) and  $H_0$ , the test statistic is

$$GQ(T_1, T_3, k) = \frac{S_3/(T_3 - k)}{S_1/(T_1 - k)} \quad (3.15)$$

---

<sup>5</sup>In a recent paper, Demos and Sentana (1998) have proposed one-sided LM tests for ARCH effects as well as critical values for LR and Wald tests which take into account the one-sided nature of the problem. Similarly, Beg, Silvapulle, and Silvapulle (1998) have introduced a one-sided sup-type generalization of the Bera-Ra test, together with simulation-based cut-off points, because of the untractable asymptotic null distributions involved. For further discussion of the difficult asymptotic distributional issues associated with such problems, see also Andrews (1999) and Klüppelberg, Maller, Van De Vyver, and Wee (2000). The MC test method should also prove to be useful with these procedures.



where  $S_1$  and  $S_3$  are the sum of squared residuals from the first  $T_1$  and the last  $T_3$  observations (where  $k < T_1$  and  $k < T_3$ ). Under the null,  $GQ(T_1, T_3, k) \sim F(T_3 - k, T_1 - k)$ . The latter distributional result is exact provided the ranking index does not depend on the parameters of the constrained model. Setting  $G_{F(T_3-k, T_1-k)}(x) = P[F(T_3 - k, T_1 - k) \geq x]$ , we denote

$$pv[GQ; T_1, T_3, k] = G_{F(T_3-k, T_1-k)}[GQ(T_1, T_3, k)] \quad (3.16)$$

the  $p$ -value associated with  $GQ(T_1, T_3, k)$ .

The  $GQ$  test is especially relevant in testing for breaks in variance.<sup>6</sup> To account for an unknown (or unspecified) break-date, we propose here (as for the BPG test) statistics of the form:

$$F_{\min}(GQ; K) \equiv 1 - \min\{pv[GQ; T_1, T_3, k] : (T_1, T_3) \in K\}, \quad (3.17)$$

$$F_{\times}(GQ; K) \equiv 1 - \prod_{(T_1, T_3) \in K} pv[GQ; T_1, T_3, k], \quad (3.18)$$

where  $K$  is any appropriate non-empty subset of

$$K(k, T) = \{(T_1, T_3) \in \mathbb{Z}^2 : k + 1 \leq T_1 \leq T - k - 1 \text{ and } k + 1 \leq T_3 \leq T - T_1\}, \quad (3.19)$$

the set of the possible subsample sizes compatible with the definition of the  $GQ$  statistic. Reasonable choices for  $K$  could be  $K = S_1(T, T_2, L_0, U_0)$  with

$$S_1(T, T_2, L_0, U_0) \equiv \{(T_1, T_3) : L_0 \leq T_1 \leq U_0 \text{ and } T_3 = T - T_1 - T_2 \geq 0\}, \quad (3.20)$$

where  $T_2$  represents the number of central observations while  $L_0$  and  $U_0$  are minimal and maximal sizes for the subsamples ( $0 \leq T_2 \leq T - 2k - 2$ ,  $L_0 \geq k + 1$ ,  $U_0 \leq T - T_2 - k - 1$ ), or

$$K = S_2(T, L_0, U_0) = \{(T_1, T_3) : L_0 \leq T_1 = T_3 \leq U_0\} \quad (3.21)$$

where  $L_0 \geq k + 1$  and  $U_0 \leq I[T/2]$ ;  $I[x]$  is the largest integer less than or equal to  $x$ . According to definition (3.20),  $\{GQ(T_1, T_3, k) : (T_1, T_3) \in K\}$  defines a set of  $GQ$  statistics, such that the number  $T_2$  of central observations is kept constant (although the sets of the central observations differ across the  $GQ$  statistics considered); with (3.21),  $\{GQ(T_1, T_3, k) : (T_1, T_3) \in K\}$  leads to  $GQ$  statistics such that  $T_1 = T_3$  (hence with different numbers of central observations). As with the BPG statistics, we also consider

$$F_{\times}(GQ; \hat{K}_{(4)}) \equiv 1 - \prod_{(T_1, T_3) \in \hat{K}_{(4)}} pv[GQ; T_1, T_3, k]$$

where  $\hat{K}_{(4)}$  selects the four smallest  $p$ -values from the set  $\{pv[GQ; T_1, T_3, k] : (T_1, T_3) \in K\}$ .

It is clear the null distribution of these statistics may be quite difficult to obtain, even asymptotically. We will see below that the level of a test procedure based on any one of these statistics can

---

<sup>6</sup>Pagan and Hall (1983, page 177) show that the  $GQ$  test for a break in variance and the relevant dummy-variable based  $BPG$  test are highly related.

be controlled quite easily on using the MC version of these tests.

### 3.3.2. Generalized Bartlett tests

Under the Gaussian assumption (2.6), the likelihood ratio criterion for testing  $H_0$  against  $H_4$  is a (strictly) monotone increasing transformation of the statistic:

$$LR_{(H_4)} = T \ln(\hat{\sigma}^2) - \sum_{i=1}^p T_i \ln(\hat{\sigma}_i^2) \quad (3.22)$$

where  $\hat{\sigma}^2$  is the ML estimator (assuming i.i.d. Gaussian errors) from the pooled regression (2.1) while  $\hat{\sigma}_i^2$ ,  $i = 1, \dots, p$ , are the ML estimators of the error variances for the  $p$  subgroups [which, due to the common regression coefficients require an iterative estimation procedure]. If one further allows the regression coefficient vectors to differ between groups (under both the null and the alternative hypothesis), one gets the extension to the linear regression setup of the well-known Bartlett (1937) test for variance homogeneity.<sup>7</sup> Note Bartlett (1937) studied the special case where the only regressor is a constant, which is allowed to differ across groups. Other (quasi-LR) variants of the Bartlett test, involving degrees of freedom corrections or different ways of estimating the group variances, have also been suggested; see, for example, Binkley (1992).

In the context of  $H_2$ , Ramsey (1969) suggested a modification to Bartlett's test that can be run on BLUS residuals from ordered observations. Following Griffiths and Surekha (1986), we consider an OLS-based version of Ramsey's test which involves separating the residuals  $\hat{u}_{(t)}$ ,  $t = 1, \dots, T$ , into three disjoint subsets  $G_i$  with  $T_i$ ,  $i = 1, 2, 3$ , observations respectively. The test statistic is:

$$RB = T \ln(\hat{\sigma}^2) - \sum_{i=1}^3 T_i \ln(\hat{\sigma}_i^2), \quad (3.23)$$

where

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_{(t)}^2, \quad \hat{\sigma}_i^2 = \frac{1}{T_i} \sum_{t \in G_i} \hat{u}_{(t)}^2.$$

Under the null,  $RB \overset{asy}{\rightsquigarrow} \chi^2(2)$ .

### 3.3.3. Szroeter-type tests

Szroeter (1978) introduced a wide class of tests. The Szroeter tests are based on statistics of the form

$$\tilde{h} = \left( \sum_{t \in A} h_t \tilde{u}_t^2 \right) / \left( \sum_{t \in A} \tilde{u}_t^2 \right) \quad (3.24)$$

<sup>7</sup>In this case, the estimated variances are  $\hat{\sigma}_i^2 = S_i/T_i$ ,  $i = 1, \dots, p$ , and  $\hat{\sigma}^2 = \sum_{i=1}^p T_i S_i/T$ , where  $S_i$  is the sum of squared errors from a regression which only involves the observations in the  $i$ -th group. This of course requires one to use groups with sufficient numbers of observations.

where  $A$  is some non-empty subset of  $\{1, 2, \dots, T\}$ , the  $\tilde{u}_t$ 's are a set of residuals, the  $h_t$ 's are a set of nonstochastic scalars such that  $h_s \leq h_t$  if  $s < t$ . Szroeter suggested several special cases of this statistic (obtained by selecting different weights  $h_t$ ), among which we shall consider the following ones which are based on the OLS residuals from a single regression  $[\tilde{u}_t = \hat{u}_{(t)}]$  :

$$SKH = \left[ \sum_{t=1}^T 2(1 - \cos \frac{\pi t}{t+1}) \hat{u}_{(t)}^2 \right] / \left( \sum_{t=1}^T \hat{u}_{(t)}^2 \right), \quad (3.25)$$

$$S_N = \left( \frac{6T}{T^2 - 1} \right)^{1/2} \left( \frac{\sum_{t=1}^T t \hat{u}_{(t)}^2}{\sum_{t=1}^T \hat{u}_{(t)}^2} - \frac{T+1}{2} \right), \quad (3.26)$$

$$S_F = \left( \sum_{t=T_1+T_2+1}^T \hat{u}_{(t)}^2 \right) / \left( \sum_{t=1}^{T_1} \hat{u}_{(t)}^2 \right) \equiv S_F(T_1, T - T_1 - T_2). \quad (3.27)$$

Under the null hypothesis,  $S_N$  follows a  $N(0, 1)$  distribution asymptotically. Exact critical points for  $SKH$  [under (2.6)] may be obtained using the Imhof method. Szroeter recommends the following bounds tests. Let  $h_L^*$  and  $h_U^*$  denote the bounds for the Durbin and Watson (1950) test corresponding to  $T + 1$  observations and  $k$  regressors. Reject the homoskedasticity hypothesis if  $SKH > 4 - h_L^*$ , accept if  $SKH < 4 - h_U^*$ , and otherwise treat the model as inconclusive. King (1981) provided revised bounds for use with  $SKH$  calculated from data sorted such that, under the alternative, the variances are non-increasing. Harrison (1980, 1981, 1982) however showed there is a high probability that the Szroeter and King bounds tests be inconclusive; in view of this, he derived and tabulated beta-approximate critical values based on the Fisher distribution.

As with the  $GQ$  test, the Szroeter's  $S_F$  statistic may be interpreted as a variant of the  $GQ$  statistic where the residuals from separate regressions have been replaced by those from the regression based on all the observations, so that  $S_3$  is replaced by  $\tilde{S}_3 = \sum_{t=T_1+T_2+1}^T \hat{u}_{(t)}^2$  and  $S_1$  by  $\tilde{S}_1 = \sum_{t=1}^{T_1} \hat{u}_{(t)}^2$ . Harrison and McCabe (1979) suggested a related test statistic based on the ratio of the sum of squares of a subset of  $\{\hat{u}_{(t)}, t = 1, \dots, T, \}$  to the total sum of squares:

$$HM = \left( \sum_{t=1}^{T_1} \hat{u}_{(t)}^2 \right) / \left( \sum_{t=1}^T \hat{u}_{(t)}^2 \right) \quad (3.28)$$

where  $T_1 = I[T/2]$ . Although the test critical points may also be derived using the Imhof method, Harrison and McCabe proposed the following bounds test. Let

$$b_L^* = \left( 1 + \frac{(T - T_1)F_\alpha(T - T_1, T - k)}{T - k} \right)^{-1}, \quad (3.29)$$

$$b_U^* = \left( 1 + \frac{(T - T_1 - k)F_\alpha(T - T_1 - k, T_1)}{T - k} \right)^{-1}, \quad (3.30)$$

where  $F_\alpha(l_1, l_2)$  refers to the level  $\alpha$  critical value from the  $F(l_1, l_2)$  distribution.  $H_0$  is rejected if  $HM < b_L^*$ , it is accepted if  $HM > b_U^*$ , and otherwise the test is inconclusive. Beta approximations

to the null distribution of the  $HM$  statistic have also been suggested. However, as argued in Harrison (1981), the two-moment beta approximation offers little savings in computational cost over the exact tests.

McCabe (1986) proposed a generalization of the  $HM$  test to the case of heteroskedasticity occurring at unknown points. The test involves computing the maximum  $HM$  criterion over several sample subgroups (of size  $m$ ). The author suggests Bonferroni-based significance points using the quantiles of the Beta distribution with parameters  $[m/2, (t - m - k)/2]$ . McCabe discusses an extension to the case where  $m$  is unknown. The proposed test is based on the maximum of the successive differences of the order statistics and also uses approximate beta critical points.

### 3.3.4. Generalized Cochran-Hartley tests

Cochran (1941) and Hartley (1950) proposed two classic tests against grouped heteroskedasticity (henceforth denoted  $C$  and  $H$ , respectively) in the context of simple Gaussian location-scale models (i.e., regressions that include only a constant). These are based on maximum and minimum subgroup error variances. Extensions of these tests to the more general framework of linear regressions have been considered by Rivest (1986). The relevant statistics then take the form:

$$C = \max_{1 \leq i \leq p} (s_i^2) / \sum_{i=1}^p s_i^2, \quad (3.31)$$

$$H = \max_{1 \leq i \leq p} (s_i^2) / \min_{1 \leq i \leq p} (s_i^2), \quad (3.32)$$

where  $s_i^2$  is the unbiased error variance estimator from the  $i$ -th separate regression ( $1 \leq i \leq p$ ). Although critical values have been tabulated for the simple location-scale model [see Pearson and Hartley (1976, pp. 202-203)], these are not valid for more general regression models, and Rivest (1986) only offers an asymptotic justification.

We will see below that the Cochran and Hartley tests can easily be implemented as finite-sample MC tests in the context of the regression model (2.1) - (2.4). Further, it will be easy to use in the same way variants of these tests that may be easier to implement or more powerful than the original procedures. Here we shall study Cochran and Hartley-type tests where the residuals from separate regressions are replaced by the OLS residuals from the pooled regression (2.1), possibly after the data have been resorted according to some exogenous variable. This will reduce the loss in degrees of freedom due to the separate regressions. The resulting test statistics will be denoted  $C_r$  and  $HR_r$  respectively. Clearly, standard distributional theory does not apply to these modified test criteria, but they satisfy the conditions required to implement them as MC tests.

### 3.3.5. Grouping tests against a mean dependent variance

Most of the tests based on grouping, as originally suggested, are valid for alternatives of the form  $H_2$ . A natural extension to alternatives such as  $H_3$  involves sorting the data conformably with  $\hat{y}_t$ ; see Pagan and Hall (1983) or Ali and Giaccotto (1984). However this complicates the finite-sample distributional theory. In this case, or whenever the alternative tested requires ordering the

sample following the fitted values of a preliminary regression, we suggest the following variants of the tests. Rather than sorting the data, sort the residuals  $\hat{u}_t$ ,  $t = 1, \dots, T$ , following  $\hat{y}_t$  and proceed. Provided the fitted values  $(\hat{y}_1, \dots, \hat{y}_T)'$  are independent of the least-squares residuals  $(\hat{u}_1, \dots, \hat{u}_T)'$  under the null hypothesis, as occurs for example under the Gaussian assumption (2.6), this will preserve the pivotal property of the tests and allow the use of MC tests.

## 4. Finite-sample distributional theory

We will now show that all the statistics described in Section 3 have null distributions which are free of nuisance parameters and show how this fact can be used to perform a finite-sample MC test of homoskedasticity using any one of these statistics. For that purpose, we shall exploit the following general proposition.

**Proposition 4.1** CHARACTERIZATION OF PIVOTAL STATISTICS. *Under the assumptions and notations (2.1) – (2.2), let  $S(y, X) = (S_1(y, X), S_2(y, X), \dots, S_m(y, X))'$  be any vector of real-valued statistics  $S_i(y, X)$ ,  $i = 1, \dots, m$ , such that*

$$S(cy + Xd, X) = S(y, X), \text{ for all } c > 0 \text{ and } d \in \mathbb{R}^k. \quad (4.1)$$

*Then, for any positive constant  $\sigma_0 > 0$ , we can write*

$$S(y, X) = S(u/\sigma_0, X), \quad (4.2)$$

*and the conditional distribution of  $S(y, X)$ , given  $X$ , is completely determined by the matrix  $X$  and the conditional distribution of  $u/\sigma_0 = \Delta\varepsilon/\sigma_0$  given  $X$ , where  $\Delta = \text{diag}(\sigma_t : t = 1, \dots, T)$ . In particular, under  $H_0$  in (2.4), we have*

$$S(y, X) = S(\varepsilon, X) \quad (4.3)$$

*where  $\varepsilon = u/\sigma$ , and the conditional distribution of  $S(y, X)$ , given  $X$ , is completely determined by the matrix  $X$  and the conditional distribution of  $\varepsilon$  given  $X$ .*

PROOF. The result follows on taking  $c = 1/\sigma_0$  and  $d = -\beta/\sigma_0$ , which entails, by (2.1),

$$cy + Xd = (X\beta + u)/\sigma_0 - X\beta/\sigma_0 = u/\sigma_0.$$

Then, using (4.1), we get (4.2), so the conditional distribution of  $S(y, X)$  only depends on  $X$  and the conditional distribution of  $u/\sigma_0$  given  $X$ . The identity  $u_0 = \Delta\varepsilon$  follows from (2.2). Finally, under  $H_0$  in (2.4), we have  $u = \Delta\varepsilon = \sigma\varepsilon$ , hence, on taking  $\sigma_0 = \sigma$ , we get  $u/\sigma_0 = \varepsilon$  and  $S(y, X) = S(\varepsilon, X)$ . ■

It is of interest to note that the representation (4.2) holds under both the general heteroskedastic model (2.1) - (2.2) and the homoskedastic model obtained by imposing (2.4), without any parametric

distributional assumption on the disturbance vector  $u$  [such as (2.3)]. Then, assuming (2.3), we see that the conditional distribution of  $S(y, X)$ , given  $X$ , is free of nuisance parameters and thus may be simulated. Of course, the same will hold for any transformation of the components of  $S(y, X)$ , such as statistics defined as the supremum or the product of several statistics (or  $p$ -values).

It is relatively easy to check that all the statistics described in Section 3 satisfy the invariance condition (4.1). In particular, on observing that model (2.1)-(2.2) and the hypothesis (2.4) are invariant to general transformations of  $y$  to  $y_* = cy + Xd$ , where  $c > 0$  and  $d \in \mathbb{R}^k$ , on  $y$ , it follows that LR test statistics against heteroskedasticity, such the Bartlett test based on  $LR_{(H_4)}$  in (3.22), satisfy (4.1) [see Dagenais and Dufour (1991) and Dufour and Dagenais (1992)], and so have null distributions which are free of nuisance parameters.

For the other statistics, the required results follow on observing that they are scale-invariant functions of OLS residuals. For that purpose, it will be useful to state the following corollary of Proposition 4.1.

**Corollary 4.2** PIVOTAL PROPERTY OF RESIDUAL-BASED STATISTICS. *Under the assumptions and notations (2.1) – (2.2), let  $S(y, X) = (S_1(y, X), S_2(y, X), \dots, S_m(y, X))'$  be any vector of real-valued statistics  $S_i(y, X)$ ,  $i = 1, \dots, m$ , such that  $S(y, X)$  can be written in the form*

$$S(y, X) = \bar{S}(A(X)y, X), \quad (4.4)$$

where  $A(X)$  is any  $n \times k$  matrix ( $n \geq 1$ ) such that

$$A(X)X = 0 \quad (4.5)$$

and  $\bar{S}(A(X)y, X)$  satisfies the scale-invariance condition

$$\bar{S}(cA(X)y, X) = \bar{S}(A(X)y, X), \text{ for all } c > 0. \quad (4.6)$$

Then, for any positive constant  $\sigma_0 > 0$ , we can write

$$S(y, X) = \bar{S}(A(X)u/\sigma_0, X) \quad (4.7)$$

and the conditional distribution of  $S(y, X)$ , given  $X$ , is completely determined by the matrix  $X$  and the conditional distribution of  $A(X)u/\sigma_0$  given  $X$ . In particular, under  $H_0$  in (2.4), we have  $S(y, X) = \bar{S}(A(X)\varepsilon, X)$ , where  $\varepsilon = u/\sigma$ , and the conditional distribution of  $S(y, X)$ , given  $X$ , is completely determined by the matrix  $X$  and the conditional distribution of  $A(X)\varepsilon$  given  $X$ .

It is easy to see that the conditions (4.4)-(4.6) are satisfied by any scale-invariant function of the OLS residuals from the regression (2.1), i.e. any statistic of the form  $S(y, X) = \bar{S}(\hat{u}, X)$  such that  $\bar{S}(c\hat{u}, X) = \bar{S}(\hat{u}, X)$  for all  $c > 0$  [in this case, we have  $A(X) = I_T - X(X'X)^{-1}X'$ ]. This applies to all the tests based on auxiliary regressions described in Section 3.1 as well as the tests against ARCH-type heteroskedasticity [Section 3.2].<sup>8</sup> On the other hand, the tests designed against

---

<sup>8</sup>For the case of the Breusch-Pagan test, the fact that the test statistic follows a null distribution free of nuisance parameters has been pointed out by Breusch and Pagan (1979) and Pagan and Pak (1993), although no proof is provided

grouped heteroskedasticity [Section 3.3] involve residuals from subsets of observations. These also satisfy the sufficient conditions of Corollary 4.2 although the  $A(X)$  matrix involved is different. For example, for the Goldfeld-Quandt statistic, we have:

$$A(X) = \begin{bmatrix} M(X_1) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & M(X_3) \end{bmatrix} \quad (4.8)$$

where  $M(X_i) = I_{T_i} - X_i(X_i'X_i)^{-1}X_i'$ ,  $X = [X_1', X_2', X_3']'$  and  $X_i$  is a  $T_i \times k$  matrix. Note the number  $n$  of rows in  $A(X)$  can be as large as one wishes so several regressions of this type can be used to compute the test statistic, as done for the combined  $GQ$  statistic [see (3.17)].

Let us now make the parametric distributional assumption (2.3). Then we can proceed as follows to perform a finite-sample test based on any statistic, say  $S_0 = S(y, X)$ , whose null distribution (given  $X$ ) is free of nuisance parameters. Let  $G(x)$  be the survival function associated with  $S_0$  under  $H_0$ , i.e. we assume  $G : R \rightarrow [0, 1]$  is the function such that  $G(x) = P_{H_0}[S_0 \geq x]$  for all  $x$ , where  $P_{H_0}$  refers to the relevant probability measure (under  $H_0$ ). Without loss of generality, we consider a right-tailed procedure:  $H_0$  rejected at level  $\alpha$  when  $S_0 \geq c(\alpha)$ , where  $c(\alpha)$  is the appropriate critical value such that  $G[c(\alpha)] = \alpha$ , or equivalently (with probability 1) when  $G(S_0) \leq \alpha$  [i.e. when the  $p$ -value associated with the observed value of the statistic is less than or equal to  $\alpha$ ].

Now suppose we can generate (by Monte Carlo methods)  $N$  i.i.d. replications of the error vector  $\varepsilon$  according to assumption (2.3). This leads to  $N$  simulated samples and  $N$  independent realizations of the test statistic  $S_1, \dots, S_N$ . The associated MC critical region obtains as

$$\hat{p}_N(S_0) \leq \alpha \quad (4.9)$$

where

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1} \quad (4.10)$$

and

$$\hat{G}_N(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{[0, \infty)}(S_i - x), \quad \mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}. \quad (4.11)$$

Then, provided the distribution function of  $S_0$  induced by  $P_{H_0}$  under  $H_0$  is continuous, we have:

$$P_{H_0}[\hat{p}_N(S_0) \leq \alpha] = \frac{I[\alpha(N + 1)]}{N + 1}, \quad \text{for } 0 \leq \alpha \leq 1, \quad (4.12)$$

see Dufour and Kiviet (1998).<sup>9</sup> In particular, if  $N$  is chosen so that  $\alpha(N + 1)$  is an integer, we have:

$$P_{H_0}[\hat{p}_N(S_0) \leq \alpha] = \alpha. \quad (4.13)$$

by them. The results given here provide a rigorous justification and considerably extend this important observation.

<sup>9</sup>If the distribution of  $S_0$  was not continuous, then we can write  $P_{H_0}[\hat{p}_N(S_0) \leq \alpha] \leq I[\alpha(N + 1)] / (N + 1)$ , so the procedure described here will be conservative. Under the assumption (2.3), however, the statistics described in section 3 have continuous null distributions.

Thus the critical region (4.9) has the same size as the critical region  $G(S_0) \leq \alpha$ .

The MC test so obtained is theoretically exact, irrespective of the number  $N$  of replications used. Note that the procedure is closely related to the parametric bootstrap, with however a fundamental difference: bootstrap tests are, in general, provably valid for  $N \rightarrow \infty$ . See Dufour and Kiviet (1996, 1998), Kiviet and Dufour (1997) and Dufour and Khalaf (2001) for some econometric applications of MC tests.

Finally, it is clear from the statement of Assumption (2.3) that normality does not play any role for the validity of the MC procedure just described. So we can consider in this way alternative error distributions such as heavy-tailed distributions like the Cauchy distribution. The only tests for which normality may play a central role in order to control size are those designed against a variance which is a function of the mean and where the least squares (LS) residuals are sorted according to the LS fitted values  $\hat{y}_t$ ,  $t = 1, \dots, T$ . Since the distribution of the latter depends on nuisance parameters (for example, the regression coefficients  $\beta$ ), it is not clear that a test statistic which depends on both  $\hat{u} = (\hat{u}_1, \dots, \hat{u}_T)'$  and  $\hat{y} = (\hat{y}_1, \dots, \hat{y}_T)'$  will have a null distribution free of nuisance parameters under the general distributional assumption (2.3). However, if we make the Gaussian assumption (2.6),  $\hat{u}$  and  $\hat{y}$  are mutually independent under  $H_0$ , and we can apply the argument justifying the MC test procedure conditional on  $\hat{y}$ .

## 5. Simulation experiments

In this section, we present simulation results illustrating the performance of the procedures described in the preceding sections. Because the number of tests and alternative models is so large, we have gathered our results in four groups corresponding to four basic types of alternatives: (1) GARCH-type heteroskedasticity; (2) variance as a linear function of exogenous variables; (3) grouped heteroskedasticity; (4) variance break at a (possibly unspecified) point.

### 5.1. Tests for ARCH and GARCH effects

For ARCH and GARCH alternatives, our simulation experiment was based on the following specification:

$$y_t = x_t' \beta + \varepsilon_t h_t^{\frac{1}{2}} + h_t \phi, \quad (5.14)$$

$$h_t = \bar{\omega}_0 + (\alpha_0 \varepsilon_{t-1}^2 + \alpha_1) h_{t-1}, \quad t = 1, \dots, T, \quad (5.15)$$

where  $x_t$  is defined as in (2.1), and  $\varepsilon_t \stackrel{iid}{\sim} N(0, 1)$ ,  $t = 1, \dots, T$ . We took  $T = 25, 50, 100$  and  $k = I[T^{1/2}] + 1$ . In the case of tests against ARCH-M alternatives [experiment (iv)], we also considered a number of alternative error distributions, according to the examples studied by Godfrey (1996):  $N(0, 1)$ ,  $\chi^2(2)$ ,  $U[-.5, .5]$ ,  $t(5)$  and *Cauchy*.

The data were generated setting  $\beta = (1, 1, \dots, 1)'$  and  $\bar{\omega}_0 = 1$ . In this framework, the model with  $\phi = 0$  and  $\alpha_1 = 0$  is a standard ARCH(1) specification, while  $\phi = 0$  yields a GARCH(1,1) model. The ARCH-M system discussed in Bera and Ra (1995) corresponds to  $\alpha_1 = 0$ . The parameter combinations considered are given in Table 2 and were selected to make the results of our



Table 2. Parameter values used for the GARCH models

Experiment	$\phi$	$\alpha_0$	$\alpha_1$
(i)	0	0	0
(ii)	0	.1, .5, .9, 1	0
(iii)	0	.1	.5
	0	.25	.65
	0	.4	.5
	0	.15	.85
	0	.05	.95
(iv)	-2, -1, 1, 2	0, .1, .5, .9	0

Table 3. Tests against ARCH and GARCH heteroskedasticity

			$T = 25$		$T = 50$		$T = 100$	
	$(\phi, \alpha_0, \alpha_1)$		Engle	Lee-King	Engle	Lee-King	Engle	Lee-King
$H_0$	(0, 0, 0)	ASY	2.21	3.22	3.06	4.04	3.63	4.72
		MC	5.01	5.14	4.94	5.01	5.18	5.22
ARCH	(0, .1, 0)	ASY	3.26	4.98	6.42	9.67	11.83	17.01
		MC	6.16	<b>7.94</b>	8.60	<b>11.31</b>	13.61	<b>17.22</b>
	(0, .5, 0)	ASY	10.19	15.73	31.56	39.96	64.18	<b>71.93</b>
		MC	14.87	<b>20.66</b>	35.68	<b>42.28</b>	66.43	71.54
	(0, .9, 0)	ASY	17.14	24.93	50.57	58.89	84.38	<b>88.99</b>
		MC	23.10	<b>31.17</b>	54.76	<b>60.71</b>	85.82	88.97
	(0, 1, 0)	ASY	24.49	26.80	53.43	61.82	86.40	<b>90.60</b>
		MC	18.59	<b>33.03</b>	57.77	<b>63.61</b>	87.50	90.24
	(0, .1, .5)	ASY	3.07	5.16	6.89	10.45	12.54	17.89
		MC	6.10	<b>8.08</b>	9.16	<b>12.05</b>	14.39	<b>18.35</b>
GARCH	(0, .25, .65)	ASY	4.87	8.72	16.26	23.43	38.36	46.93
		MC	8.44	<b>12.60</b>	19.92	<b>25.20</b>	40.74	<b>47.29</b>
	(0, .40, .50)	ASY	7.30	12.66	26.12	34.48	57.44	65.94
		MC	11.89	<b>17.07</b>	30.25	<b>36.51</b>	59.65	<b>65.69</b>
	(0, .15, .85)	ASY	4.42	8.40	13.45	19.96	28.97	37.11
		MC	7.93	<b>12.39</b>	16.79	<b>21.70</b>	31.04	<b>37.93</b>
	(0, .05, .95)	ASY	3.75	7.31	10.02	15.37	18.17	25.92
		MC	6.84	<b>10.74</b>	12.77	<b>17.05</b>	20.15	<b>26.28</b>

Note: In this table as well as in the subsequent ones, the rejection frequencies are reported in percentages.

Table 4. Empirical size of ARCH-M tests

Parameter values			Error distribution									
$\alpha_0 = \alpha_1 = 0, \phi = -2$			$N(0, 1)$		$\chi^2(2)$		$U[-.5, .5]$		$t(5)$		<i>Cauchy</i>	
Test		$T$	D1	D2	D1	D2	D1	D2	D1	D2	D1	D2
Engle	ASY	25	2.2	2.2	1.8	1.3	2.7	2.4	2.3	0.8	1.8	0.6
	MC		5.6	5.2	4.3	4.0	5.1	4.8	5.3	4.3	5.0	4.9
Lee-King	ASY	25	3.4	4.2	2.8	2.3	2.7	3.7	2.4	1.7	1.9	1.4
	MC		5.0	4.8	5.4	4.0	4.6	4.5	5.1	4.0	5.2	5.3
Bera-Ra	MC	25	4.7	4.5	3.6	4.1	5.4	4.6	4.9	4.7	5.2	5.5
Engle	ASY	50	2.6	2.9	3.2	3.0	4.7	3.4	2.6	2.6	1.6	1.7
	MC		4.6	4.2	6.3	6.2	6.5	4.5	6.2	5.0	5.8	5.5
Lee-King	ASY	50	4.3	4.6	4.1	4.4	5.1	6.1	3.7	3.6	2.3	2.3
	MC		5.6	4.0	5.0	5.5	5.7	4.5	5.7	5.7	5.5	5.0
Bera-Ra	MC	50	5.0	4.8	5.6	5.1	5.1	4.9	5.0	4.9	4.8	4.8
Engle	ASY	100	4.1	3.9	2.9	2.9	5.2	4.8	2.7	2.8	1.8	1.8
	MC		4.7	5.3	5.2	5.2	5.6	5.4	5.3	4.8	5.9	5.2
Lee-King	ASY	100	5.3	5.4	4.2	4.5	4.1	6.0	3.7	3.4	2.4	2.3
	MC		5.3	5.1	5.4	5.3	4.5	5.0	5.4	5.0	5.5	5.2
Bera-Ra	MC	100	5.6	6.2	5.2	5.2	5.1	4.9	5.1	4.5	4.7	5.3

Table 5. Power of MC ARCH-M tests: normal errors and D1 design

$\alpha_0$	$\phi$	Test	$T = 25$			$T = 50$			$T = 100$		
			Engle	Lee-King	Bera-Ra	Engle	Lee-King	Bera-Ra	Engle	Lee-King	Bera-Ra
.1	-2	ASY	3.6	5.7		7.7	9.5		16.0	21.2	
		MC	6.3	9.2	<b>9.4</b>	9.5	<b>13.9</b>	12.6	18.5	21.5	<b>24.5</b>
.1	-1	ASY	3.8	5.7		6.7	9.8		14.2	18.6	
		MC	6.4	<b>8.9</b>	8.7	8.4	<b>11.9</b>	9.8	16.0	<b>18.2</b>	16.4
.1	1	ASY	3.8	5.4		6.9	9.7		13.5	17.8	
		MC	5.9	7.6	<b>8.3</b>	8.2	10.9	<b>10.10</b>	15.1	<b>17.1</b>	<b>17.1</b>
.1	2	ASY	3.8	5.4		7.8	11.3		15.4	21.0	
		MC	6.4	8.4	<b>9.0</b>	9.8	12.5	<b>13.0</b>	17.5	20.6	<b>23.9</b>
.5	-2	ASY	10.5	16.7		31.4	41.7		67.2	73.0	
		MC	15.6	21.9	<b>26.4</b>	36.5	43.6	<b>56.2</b>	69.1	72.7	<b>84.5</b>
.5	-1	ASY	9.9	17.1		33.5	42.9		68.3	76.3	
		MC	16.2	<b>22.9</b>	22.6	38.5	45.2	<b>47.5</b>	70.5	75.6	<b>78.1</b>
.5	1	ASY	10.2	15.2		32.4	41.5		64.8	74.8	
		MC	14.4	19.6	<b>21.0</b>	36.8	43.2	<b>46.2</b>	67.0	73.4	<b>77.1</b>
.5	2	ASY	10.4	16.6		34.4	43.8		63.3	72.5	
		MC	16.0	22.4	<b>26.6</b>	38.0	45.3	<b>56.6</b>	66.3	70.9	<b>84.6</b>
.9	-2	ASY	13.3	20.9		41.0	49.7		67.6	72.8	
		MC	19.6	24.7	<b>33.6</b>	45.6	50.6	<b>63.6</b>	69.4	72.6	<b>86.6</b>
.9	-1	ASY	16.4	24.6		46.7	56.4		79.3	84.2	
		MC	22.6	30.6	<b>33.0</b>	52.1	58.8	<b>67.5</b>	80.7	84.1	<b>91.6</b>
.9	1	ASY	15.7	23.7		47.0	56.6		75.7	81.5	
		MC	21.3	29.5	<b>31.9</b>	51.2	58.1	<b>65.3</b>	77.6	82.0	<b>89.9</b>
.9	2	ASY	14.9	22.9		40.7	48.3		65.9	70.4	
		MC	21.2	27.0	<b>33.3</b>	44.3	49.8	<b>63.5</b>	66.9	70.1	<b>85.1</b>

Table 6. Power of MC ARCH-M tests: various error distributions and D2 design

$\phi = -2$			$T = 25$			$T = 50$			$T = 100$		
$\alpha_0$	Error	Test	Engle	Lee-King	Bera-Ra	Engle	Lee-King	Bera-Ra	Engle	Lee-King	Bera-Ra
.1	$N(0, 1)$	ASY	4.1	7.5		8.4	13.2		17.4	24.8	
		MC	8.2	8.8	<b>9.3</b>	11.2	12.5	<b>13.8</b>	19.8	23.3	<b>25.0</b>
	$\chi^2(2)$	ASY	9.3	16.4		30.1	38.9		54.5	61.3	
		MC	20.7	23.0	<b>27.6</b>	40.2	42.2	<b>47.5</b>	59.7	60.5	<b>65.4</b>
	$U[-.5, .5]$	ASY	2.3	4.0		3.2	6.0		5.1	7.0	
		MC	4.7	4.1	<b>4.6</b>	4.4	4.9	<b>5.0</b>	5.6	<b>6.2</b>	5.2
	$t(5)$	ASY	4.5	9.4		22.4	34.6		53.2	61.4	
		MC	20.9	26.7	<b>37.1</b>	47.3	51.4	<b>60.9</b>	71.1	72.6	<b>79.0</b>
	<i>Cauchy</i>	ASY	12.5	21.3		45.0	52.4		66.1	71.3	
		MC	32.6	36.9	<b>41.6</b>	64.0	67.8	<b>73.7</b>	81.7	83.9	<b>86.7</b>
.5	$N(0, 1)$	ASY	12.2	22.5		33.1	42.2		61.8	69.8	
		MC	20.6	24.3	<b>32.8</b>	36.7	41.0	<b>55.7</b>	64.6	68.3	<b>81.4</b>
	$\chi^2(2)$	ASY	18.6	22.4		33.3	42.3		52.9	59.8	
		MC	25.4	26.9	<b>35.4</b>	42.9	43.8	<b>61.7</b>	58.3	58.3	<b>77.4</b>
	$U[-.5, .5]$	ASY	2.3	4.6		4.3	7.6		7.4	13.0	
		MC	4.8	5.4	<b>6.6</b>	5.1	6.1	<b>8.0</b>	7.5	10.5	<b>15.4</b>
	$t(5)$	ASY	12.1	17.9		50.7	58.5		70.2	74.6	
		MC	30.1	34.0	<b>38.9</b>	67.2	69.1	<b>71.6</b>	80.5	81.3	<b>84.4</b>
	<i>Cauchy</i>	ASY	24.0	30.9		64.1	70.7		78.4	82.3	
		MC	40.2	47.7	<b>51.9</b>	77.7	81.6	<b>83.8</b>	89.3	91.1	<b>92.3</b>
.9	$N(0, 1)$	ASY	9.1	18.4		37.9	46.8		61.7	68.5	
		MC	16.4	21.6	<b>27.7</b>	42.0	44.9	<b>62.6</b>	64.2	65.9	<b>82.1</b>
	$\chi^2(2)$	ASY	14.2	21.3		27.8	32.4		40.9	45.9	
		MC	24.0	27.1	<b>37.3</b>	33.9	35.0	<b>54.4</b>	43.8	44.3	<b>66.9</b>
	$U[-.5, .5]$	ASY	2.6	5.6		4.6	10.4		12.4	19.2	
		MC	5.0	6.4	<b>9.1</b>	5.6	7.4	<b>15.4</b>	11.8	17.5	<b>35.4</b>
	$t(5)$	ASY	16.8	30.1		27.5	37.6		45.8	54.2	
		MC	41.6	44.9	<b>50.2</b>	47.8	50.9	<b>57.7</b>	62.3	63.5	<b>73.1</b>
	<i>Cauchy</i>	ASY	21.6	32.1		54.4	62.3		71.6	76.5	
		MC	44.5	50.1	<b>54.6</b>	73.8	76.9	<b>80.6</b>	85.2	87.5	<b>90.0</b>

study comparable with those obtained by Lee and King (1993), Lumsdaine (1995) and Bera and Ra (1995). Note that some combinations fall on the boundary of the region  $\alpha_0 + \alpha_1 \leq 1$ .

The regressors were generated as i.i.d. according to a  $U(0, 10)$  distribution and kept constant over each individual experiment. In the case of tests against ARCH-M alternatives [experiment (iv)], we also considered an alternative regressor set, obtained by drawing (independently) from a Cauchy distribution (centered and re-scaled conformably with the previous design). For further reference, we shall denote by D1 the *uniform*-based design D1 and by D2 the *Cauchy*-based design. Both D1 and D2 include a constant regressor.

The MC tests were implemented with  $N = 99$  replications. The Engle and Lee-King tests were applied in all cases. In experiment (iv), we also applied the Bera-Ra sup-LM test (see Section 3.2 for formulae and related references), in which case we have only computed MC versions of the tests.

Tables 3, 4, 5 and 6 report rejection percentages for a nominal level of 5%; 10000 replications were considered for experiments (i) - (ii) and 1000 for experiment (iv); in these tables (as well as later ones), the figures associated with best performing exact procedures in terms of power (under the alternatives) are set in **bold** face characters.

In general, the most notable observation is that the Engle test is undersized, even with  $T = 100$ , which can lead to substantial power losses. This is in accordance with the results of Lee and King (1993) and several references cited there. Although undersize problems are evident under D1 and normal errors, more serious size distortions are observed with  $\chi^2(2)$ ,  $t(5)$  and Cauchy errors. The size of the Lee-King test is better than that of the Engle test but is still below the nominal level particularly with  $\chi^2(2)$ ,  $t(5)$  and Cauchy errors, and for  $T < 100$  under D1 and normal errors.

So MC tests yield important effective power gains. The effective power advantage of MC tests is noticeable, even with uniform designs and normal errors. In the case of  $\chi^2(2)$ ,  $t(5)$  and Cauchy errors, improvements in power are quite substantial (such as a 50% increase with  $T \leq 50$ ). The Lee-King MC test is always more powerful than the Engle test. It is also worth noting that possible problems at boundary parameter values were not observed. Further the MC Lee-King test has the best power. As emphasized in Bera and Ra (1995), power improvements in the case of ARCH and GARCH tests are especially important since failing to detect conditional variance misspecifications leads to inconsistencies in conditional moments estimates.

Turning now to ARCH-M tests, we see that: (i) the Engle test has the lowest power among the three tests considered, and (ii) there is a substantial power gain from using the sup-LM MC test. The Lee-King MC test performs better than the Engle test and in a few cases marginally outperforms the sup test. This suggests that a MC version of the Beg, Silvapulle, and Silvapulle (1998) one-sided test for ARCH-M may also result in power improvements. The power advantage of the MC sup-LM along with the documented difficulties regarding the Davies sup-LM test, makes the MC Bera-Ra test quite attractive. Further, these results show clearly that the MC test provides a straightforward finite-sample solution to the problem of unidentified nuisance parameters.

## 5.2. Tests of variance as a linear function of exogenous variables

The model used is (2.1) with  $T = 25, 50, 100$  and  $k = 6$ .<sup>10</sup> The regression coefficients were set to one. The following specifications for the error variance were considered:

(i)  $\sigma_t^2 = x_t' \alpha$ ,  $t = 1, \dots, T$ , where

$$\begin{aligned}\alpha &= (1, 0, \dots, 0)', \quad \text{under } H_0, \\ &= (1, 1, \dots, 1)', \quad \text{under } H_A.\end{aligned}$$

(ii)  $\sigma_t^2 = \alpha_0 + \alpha_1 x_{2t}$ ,  $t = 1, \dots, T$ , where

$$\begin{aligned}\alpha_0 &= 1, \quad \alpha_1 = 0, \quad \text{under } H_0, \\ \alpha_0 &= 0, \quad \alpha_1 = 1, \quad \text{under } H_A.\end{aligned}$$

The former specification implies that the variance is a linear function of  $E(Y_t)$  and the latter is the case where the variance is proportional to one regressor. The regressors are generated as  $U(0, 10)$ . The tests examined (refer to Section 3 for formulae and related references) are the following:

1. the Goldfeld-Quandt ( $GQ$ ) test [see (3.15)], with  $T_2 = T/5$ ,  $T_1 = T_3 = (T - T_2)/2$ ;
2. the Breusch-Pagan-Godfrey ( $BPG$ ) test [see (3.1)], based on the asymptotic distribution (ASY) or using the size correction formula (BRT) proposed by Honda (1988, section 2);
3. Koenker's ( $K$ ) test [see (3.2)];
4. White's ( $W$ ) test [see (3.1)];
5. Glejser's ( $G$ ) test based on (3.3);
6. Ramsey's version of Bartlett's test ( $RB$ ) [see (3.23)], with  $T_1 = T_3 = I[T/3]$  and  $T_2 = T - (T_1 + T_3)$ ;
7. Szroeter's  $S_F$  test [see (3.27)], where for convenience,  $T_1$  and  $T_2$  are set as in the  $GQ$  test;
8. Szroeter's  $SKH$  test [see (3.25)]; the critical points for the bounds test are taken from King (1981, Table 2) and the beta-approximate cut-off points from Harrison (1982, Table 4);
9. Szroeter's  $S_N$  test [see (3.26)];
10. the Harrison-McCabe ( $HM$ ) test [see (3.28)], with  $T_1 = I[T/2]$ .

In the case of (ii), we have observed that sorting the observations or the OLS residuals by the value of  $\hat{y}$  leads to equivalent MC tests. Tables 7 and 8 report rejection percentages for a nominal level of 5% and 10000 replications. The MC tests are implemented with 99 simulated samples. Based on these two experiments, we make the following observations.

---

<sup>10</sup>Tables of critical points required for the Szroeter's tests are available for  $n \leq 100$  and  $k \leq 6$ ; see King (1981) and Harrison (1982).

Table 7. Variance proportional to a regressor

		$T = 25$		$T = 50$		$T = 100$	
		$H_0$	$H_A$	$H_0$	$H_A$	$H_0$	$H_A$
Goldfeld-Quandt ( $GQ$ )	$F$	4.95	27.64	4.68	81.41	4.95	98.25
Breusch-Pagan-Godfrey ( $BPG$ )	ASY	4.13	33.84	4.14	80.57	4.59	98.75
	BRT	4.42	35.17	4.64	81.67	4.71	98.79
	MC	5.37	36.80	4.99	80.86	4.58	98.36
Koenker ( $K$ )	ASY	5.18	30.81	4.74	75.14	4.51	97.52
	MC	5.29	30.20	4.98	74.70	4.46	96.77
White ( $W$ )	ASY	0.00	0.00	2.60	20.20	4.42	34.64
	MC	5.31	8.08	4.67	26.70	4.65	33.99
Glejser ( $G$ )	$ASY_F$	5.54	34.58	5.09	80.04	4.66	98.82
	$ASY_W$	6.72	38.62	5.76	81.30	5.03	98.90
	MC	5.07	32.32	5.12	78.48	4.58	98.44
Ramsey-Bartlett ( $RB$ )	ASY	6.67	31.87	5.50	80.06	5.22	97.96
	MC	5.27	27.63	4.58	77.03	4.76	97.49
Szroeter ( $S_F$ )	MC	4.98	51.26	4.77	88.71	4.88	99.12
Szroeter ( $S_N$ )	ASY	3.11	45.46	4.28	91.94	4.87	99.63
	MC	4.97	53.45	5.08	<b>92.09</b>	4.69	<b>99.51</b>
Szroeter-King-Harisson ( $SKH$ )	$Beta$	4.01	50.34	6.41	94.71	8.32	99.83
	Bound	0.18	10.43	0.71	74.74	1.54	98.09
	Inc. bound	34.61	82.53	19.6	24.23	12.38	1.83
	MC	4.98	<b>53.87</b>	4.98	91.68	4.79	99.43
Harisson-McCabe ( $HM$ )	Bound	0.05	3.52	0.79	61.31	1.91	94.61
	Inc. bound	20.37	75.29	13.48	33.75	9.67	4.52
	MC	4.69	46.63	4.78	84.64	5.20	97.38

Table 8. Variance as a function of the mean

		$T = 25$		$T = 50$		$T = 100$	
		$H_0$	$H_A$	$H_0$	$H_A$	$H_0$	$H_A$
Goldfeld-Quandt ( $GQ$ )	$F$	4.64	5.93	5.24	11.56	4.95	22.90
Breusch-Pagan-Godfrey ( $BPG$ )	ASY	4.12	5.56	4.38	8.69	5.01	16.57
	BRT	3.28	4.70	4.19	8.39	4.95	16.22
	MC	5.06	6.81	4.74	9.54	5.02	16.30
Koenker ( $K$ )	ASY	3.34	3.96	4.02	7.06	4.32	13.77
	MC	5.17	6.13	5.08	8.10	4.73	14.63
White ( $W$ )	ASY	0.00	0.00	2.60	3.45	4.42	7.53
	MC	5.31	5.95	4.67	5.98	4.65	7.93
Glejser ( $G$ )	$ASY_F$	5.98	6.92	5.46	9.04	5.14	15.21
	$ASY_W$	11.75	12.88	7.66	12.05	5.98	17.42
	MC	4.93	5.99	5.11	8.21	5.02	14.54
Ramsey-Bartlett ( $RB$ )	ASY	5.94	7.57	5.75	11.99	5.41	21.01
	MC	5.03	6.03	5.03	11.02	4.90	19.78
Szroeter ( $S_F$ )	MC	4.66	11.18	5.18	19.29	5.3	33.61
Szroeter ( $S_N$ )	ASY	3.93	10.74	5.32	21.91	4.91	39.64
	MC	4.68	<b>12.09</b>	5.07	<b>21.38</b>	4.84	<b>38.90</b>
Szroeter-King-Harisson ( $SKH$ )	Beta	4.67	12.06	7.96	28.07	8.28	48.09
	Bound	0.19	0.82	0.96	6.17	1.50	21.02
	Bound inconc.	39.29	58.58	20.89	48.02	12.51	39.41
	MC	4.62	11.77	4.97	21.04	5.01	37.49
Harisson-McCabe ( $HM$ )	Bound	0.08	0.17	0.74	4.18	1.85	16.25
	Bound inconc.	22.29	37.47	14.67	34.61	10.06	30.44
	MC	4.98	10.69	5.02	17.82	5.48	29.25



### 5.2.1. Level

The  $BPG$ ,  $K$ ,  $S_N$  and  $W$  tests reject the null less frequently than implied by their nominal size, particularly in small samples. The  $G$  Wald-type test and the Harrison approximate  $SKH$  test have a tendency to over-reject. The bounds tests based on the  $HM$  and  $SKH$  statistics are inconclusive in a large proportion of cases. As expected, MC tests have the correct size. In the case of the  $BPG$  criterion, Honda's size correction improves both the reliability and the power properties of the test; the superiority of the MC technique is especially notable with small samples. It is worth emphasizing that, whereas Honda's formula is generally effective, it is based on an asymptotic approximation; the MC test method is theoretically exact in finite samples.

### 5.2.2. Power

Here, we again discuss the performance of the MC versions of the tests considered, because we restrict attention to size correct procedures. We observe that the  $S_N$  and the  $SKH$  MC tests (whose performance is very similar) are most powerful, followed closely by the  $S_F$  and the  $HM$  MC test, and by the  $G$  and  $BPG$  MC tests. The  $GQ$  and  $RB$  MC tests rank next whereas the  $W$  test performs very poorly. Note that the Szroeter  $GQ$ -type test  $S_F$  performs much better than the standard  $GQ$ ; this is expected since the latter is based on residuals from a single regression on the whole sample. Overall, the most noticeable fact is the superiority of the Szroeter MC tests when compared to the commonly used procedures (e.g. the Breusch-Pagan  $TR^2$  type tests). As mentioned earlier, the Szroeter tests as initially proposed have not gained popularity due to their non-standard null distributions. Given the ease with which exact MC versions of these tests can be computed, this experiment clearly demonstrates that a sizable improvement in power results from replacing the commonly used LM-type tests with either Szroeter-type MC tests. Similar conclusions are reported in Griffiths and Surekha (1986) with respect to  $S_N$ , the member of the Szroeter family whose null distribution is asymptotically normal. However, these authors also document the asymptotic tests' incorrect finite sample size.

## 5.3. Grouped heteroskedasticity

To illustrate the performance of MC tests for grouped heteroskedasticity ( $H_3$ ), we follow the design of Binkley (1992). The model used is (2.1) with

$$T_i = 15, 25, 50; \quad k_i = 4, 6, 8; \quad m = 2, 3, 4.$$

The regressors were drawn from a  $U(0, 10)$  distribution and differed across subgroups.<sup>11</sup> The regressors were drawn only once. The regression coefficients were set to one, and the variances across groups were selected so that

$$\delta = \frac{\sigma_{\max}^2}{\sigma_{\min}^2} = 1, 2, 3, 5,$$

---

<sup>11</sup>We considered other choices for the design matrices, including Cauchy, lognormal, and identical regressors (across subgroups) and obtained qualitatively similar results.

with the intermediate variances set at equal intervals, where  $\sigma_{\min}^2$  and  $\sigma_{\max}^2$  represent respectively the smallest and largest error variance among the  $m$  groups. The errors were drawn from the normal distribution. We considered the LR statistic, the Goldfeld-Quandt statistic, the Breusch-Pagan and Koenker statistics, the Glejser and White statistics, and the Cochran and Hartley criteria ( $C$ ,  $H$ ,  $C_r$ , and  $H_r$ ). We also studied alternative likelihood-based test criteria introduced in Binkley (1992, Page 565), namely  $LR1$ ,  $LR2$ ,  $LR3$  and  $BPG_2$ ,<sup>12</sup> and considered as well a Koenker-type adjustment to  $BPG_2$  (which we denote  $K_u$ ). The results are summarized in Table 9.<sup>13</sup> We report empirical rejections for a nominal size of 5% in 10000 replications. The MC test is obtained with 99 simulated samples. Our findings can be summarized as follows.

### 5.3.1. Level

In general, LM-type asymptotic tests are undersized, whereas the asymptotic LR-type tests tend to over-reject. The variants of the LM and LR tests based on residuals from individual regressions are over-sized. As expected, size problems are more severe with small samples. Although the behavior of the size-corrected  $BPG$  appears to be satisfactory, the modification technique yields over-rejections when applied to  $BPG_u$ . Note that Honda (1988) verified that the formulae does not work well in the case of the Koenker test. Finally, the empirical size of the Cochran and Hartley statistics exceeds the nominal size. In contrast, the MC versions of all the tests considered achieve perfect size control in all cases.

### 5.3.2. Power

In order to compare tests of equal size, we only discuss the power of the MC tests. First of all, we observe that the MC technique improves the effective power of the LM and White tests. Although the correction from Honda (1988) achieves a comparable effect, its application is restricted to the standard  $BPG$  criterion. Secondly, comparing the LR and QLR tests, there is apparently no advantage to using full maximum likelihood estimation [for a similar observation in the context of SURE models, see Dufour and Khalaf (1998, 1999)].

In general, the tests may be ranked in terms of power as follows.  $LR$ ,  $QLR$ ,  $BPG$  and  $H$  performed best, followed quite closely by the  $G$ ,  $K$  and  $C$ . The  $W$  test performed poorly: its power hardly exceeds the size. Overall, with the exception of the  $W$  test, no test is uniformly dominated.

The MC tests constructed using variance estimates from separate regressions have a slight power disadvantage. This is somewhat expected, since the simulated samples were drawn imposing equality of the individual regression coefficients.

Finally, note that the MC Hartley's test compared favorably with the LM and LR test. This, together with the fact that it is computationally so simple, suggest that the application of the MC

<sup>12</sup> $LR1$  is obtained as in (3.22) replacing  $s_i^2$  by estimates of group variances from partitioning  $s^2$ .  $LR2$  is obtained as in (3.22) replacing  $s_i^2$  by variance estimates from separate regressions, over the sample subgroups, and  $s^2$  by a weighted average of these.  $LR3$  is obtained like  $LR2$ , using unbiased variance estimates.  $BPG_2$  is a variant of the  $BPG$  test for  $H_3$  based on residuals from individual group regressions.

<sup>13</sup>For convenience, our notation differs from Binkley (1992). The  $QLR$  test refers to Binkley's  $LR1$ , the  $LR_u$  ( $ASY1$ ) and  $LR_u$  ( $ASY2$ ) refer to  $LR2$  and  $LR3$  tests;  $BPG_u$  corresponds to  $BPG_2$ . Note that  $LR3$  obtains as a monotonic transformation of  $LR2$ , which yields the same MC test.

Table 9. Grouped heteroskedasticity

Two Groups ( $m = 2$ )

$\delta$	$T_i$	$GQ$	$LR$		$QLR$		$LR_u$			$H_r$	$C_r$	
		$F$	ASY	MC	ASY	MC	ASY <sub>1</sub>	ASY <sub>2</sub>	MC	MC	MC	
1	15	5.17	10.10	4.92	5.52	5.01	9.88	5.67	4.94	5.01	5.01	
	25	4.95	9.08	4.79	5.14	4.97	9.00	5.23	5.02	4.97	4.97	
	50	5.18	7.61	5.35	5.26	5.25	7.62	5.26	5.30	5.25	5.25	
2	15	18.08	29.95	18.43	20.00	<b>18.65</b>	28.37	19.30	17.45	<b>18.65</b>	<b>18.65</b>	
	25	<b>31.08</b>	42.25	30.67	31.86	30.44	41.80	32.22	30.28	30.44	30.44	
	50	<b>60.04</b>	67.50	59.21	60.68	58.77	66.72	60.52	58.82	58.77	58.77	
3	15	40.16	55.80	<b>40.63</b>	52.10	39.60	54.00	42.11	39.16	39.60	39.60	
	25	<b>64.79</b>	75.56	63.44	63.93	61.81	74.50	65.84	62.51	61.81	61.81	
	50	<b>93.86</b>	96.06	93.52	93.99	93.14	95.86	94.01	93.21	93.14	93.14	
5	15	71.95	83.45	<b>71.98</b>	71.55	69.00	81.61	73.29	70.24	69.00	69.00	
	25	<b>92.88</b>	96.29	92.39	91.26	90.04	95.85	93.08	92.06	90.04	90.04	
	50	<b>99.89</b>	99.94	99.82	99.90	99.81	99.92	99.89	99.85	99.81	99.81	
		$BPG$			$BPG_u$			$K$		$K_u$		
		ASY	$BRT$	MC	ASY	$BRT$	MC	ASY	MC	ASY	MC	
		15	3.93	4.87	5.01	7.80	9.03	4.94	4.54	5.07	8.28	4.97
	1	25	4.26	4.80	4.97	7.71	8.46	5.02	4.67	5.14	8.16	5.06
		50	4.90	5.12	5.25	7.05	7.38	5.30	4.90	5.24	7.33	5.25
		2	15	15.95	18.47	<b>18.65</b>	24.27	26.81	17.45	14.70	15.73	21.83
	25		29.04	30.70	30.44	39.02	40.69	30.28	26.79	27.38	36.22	26.80
	50		58.90	60.08	58.77	65.20	66.09	58.82	56.55	55.54	63.61	55.30
	3	15	35.81	39.80	39.60	48.69	51.88	39.16	30.57	31.68	41.77	29.83
		25	60.63	62.61	61.81	71.91	73.59	62.51	55.68	55.27	66.63	54.62
		50	93.48	93.77	93.14	95.51	95.71	93.21	91.80	90.80	94.40	90.89
	5	15	66.17	69.62	69.00	78.36	80.36	70.24	54.20	55.16	67.27	52.57
25		89.92	90.67	90.04	95.02	95.53	92.06	83.59	82.67	91.54	83.12	
50		99.86	99.90	99.81	99.92	99.92	99.85	99.69	99.52	99.88	99.64	
		Glejser			White		$C$		$H$			
		ASY <sub>F</sub>	ASY <sub>W</sub>	MC	ASY	MC	ASY	MC	ASY	MC		
	1	15	5.01	6.24	5.04	3.86	5.01	10.41	4.94	9.45	4.94	
		25	4.89	5.52	5.03	4.02	4.72	6.80	5.02	8.08	5.02	
		50	5.02	5.22	5.11	4.47	5.08	4.95	5.30	6.13	5.30	
	2	15	16.38	18.74	16.09	5.20	6.28	29.52	17.45	27.46	17.45	
		25	26.33	28.22	25.94	4.55	5.17	36.79	30.28	39.92	30.28	
		50	53.56	54.76	51.87	3.86	4.26	59.38	58.82	62.78	58.82	
	3	15	32.94	36.38	31.85	6.07	7.27	55.02	39.16	52.86	39.16	
		25	54.56	56.61	53.05	4.97	5.74	69.85	62.51	72.75	62.51	
		50	89.34	89.79	88.32	3.51	4.14	93.69	93.21	94.81	93.21	
	5	15	58.09	62.03	56.38	6.97	8.40	82.42	70.24	81.00	70.24	
25		83.88	85.17	82.52	5.18	6.16	94.49	92.06	95.20	92.06		
50		99.55	99.57	99.39	3.29	3.94	99.89	99.85	99.90	99.85		

Table 9. Grouped heteroskedasticity (continued)

Three Groups ( $m = 3$ )

$\delta$	$T_i$	$GQ$	$LR$		$QLR$		$LR_u$			$C_r$	$H_r$	
		$F$	ASY	MC	ASY	MC	ASY <sub>1</sub>	ASY <sub>2</sub>	MC	MC	MC	
1	15	-	9.94	5.09	5.35	5.05	12.07	5.63	4.97	5.25	5.24	
	25	-	8.89	5.22	5.49	5.28	10.89	5.36	4.97	5.12	5.14	
	50	-	7.28	5.11	5.33	5.25	8.16	5.22	4.90	5.13	5.24	
2	15	-	25.87	15.80	17.12	<b>16.05</b>	27.79	16.00	14.75	14.58	15.75	
	25	-	35.94	<b>25.70</b>	26.99	25.66	37.63	25.43	23.94	21.89	25.43	
	50	-	59.72	<b>51.93</b>	53.45	51.84	59.98	50.84	48.99	41.26	51.37	
3	15	-	49.17	<b>35.08</b>	36.69	34.30	49.21	33.91	31.17	26.96	30.87	
	25	-	68.62	<b>57.51</b>	58.91	56.61	68.32	55.62	52.95	42.27	56.65	
	50	-	93.84	<b>90.85</b>	91.44	90.43	93.27	89.96	88.75	73.11	90.79	
5	15	-	79.44	<b>66.26</b>	66.79	63.75	78.50	64.49	60.39	43.54	64.18	
	25	-	94.63	<b>89.79</b>	89.92	88.40	94.58	89.10	86.95	65.18	88.96	
	50	-	99.90	<b>99.80</b>	99.81	99.72	99.92	99.79	99.79	93.36	99.77	
1	15	$BPG$			$BPG_u$			$K$		$K_u$		
		ASY	$BRT$	MC	ASY	$BRT$	MC	ASY	MC	ASY	MC	
		3.92	5.06	5.51	8.99	11.07	5.18	4.08	5.09	8.86	4.95	
	25	4.45	5.19	5.20	8.89	10.02	4.90	4.57	5.17	8.54	4.88	
		50	4.62	4.93	5.16	7.55	8.15	4.93	4.72	5.29	7.39	5.06
		2	15	13.15	16.09	15.68	22.12	25.60	14.07	11.74	13.46	19.11
	25		23.33	25.26	24.67	33.33	35.61	22.87	20.52	22.23	30.31	20.35
	50		50.29	51.46	50.57	56.79	58.21	47.45	47.48	47.56	54.27	45.03
	3	15	27.89	32.34	31.63	39.85	44.56	28.20	22.65	25.59	33.31	22.05
		25	52.04	54.77	53.80	62.81	65.17	48.98	45.09	46.60	56.01	41.96
		50	89.57	90.05	88.98	91.49	92.08	87.05	86.49	85.94	89.54	83.38
	5	15	53.62	58.94	57.34	67.09	71.61	50.65	41.01	44.50	54.26	38.42
		25	84.94	86.66	84.94	91.22	92.18	81.94	75.50	76.10	83.79	70.39
		50	99.72	99.74	99.58	99.85	99.85	99.54	99.37	99.11	99.65	98.82
2	15	Glejser			White			$C$		$H$		
		ASY <sub>F</sub>	ASY <sub>W</sub>	MC	ASY	MC		ASY	MC	ASY	MC	
		4.97	5.92	5.06	4.31	4.74		11.6	5.16	11.2	5.04	
	25	5.03	5.70	5.15	5.11	5.19		7.28	5.0	9.36	4.88	
		50	5.09	5.47	4.99	5.54	5.29		4.44	5.01	11.50	5.13
		2	15	13.35	15.63	13.40	5.46	5.91		24.39	13.36	25.91
	25		21.57	23.48	21.48	7.06	7.09		25.96	20.20	34.15	23.78
	50		45.46	46.58	44.61	4.87	4.46		39.49	39.47	66.72	48.72
	3	15	27.21	30.66	26.89	5.93	6.28		39.14	24.78	47.65	30.87
		25	47.76	50.06	46.33	7.88	7.90		46.38	39.00	65.58	52.89
		50	85.19	85.84	83.87	4.40	4.12		70.66	70.31	95.41	89.01
	5	15	50.97	55.47	49.40	6.36	6.78		56.27	39.89	77.55	60.94
		25	80.72	82.35	78.74	8.76	8.73		68.83	60.83	93.78	87.53
		50	99.29	99.34	99.06	3.86	3.68		92.24	91.66	99.94	99.77

Table 9. Grouped heteroskedasticity (continued)

Four Groups ( $m = 4$ )

$\delta$	$T_i$	$GQ$	$LR$		$QLR$		$LR_u$			$C_r$	$H_r$
		$F$	ASY	MC	ASY	MC	ASY <sub>1</sub>	ASY <sub>2</sub>	MC	MC	MC
1	15	-	9.70	5.25	5.7	5.3	14.41	5.75	5.34	5.60	5.30
	25	-	8.05	4.64	5.23	4.92	11.82	5.35	5.01	4.96	4.94
	50	-	6.70	5.05	5.24	4.90	9.02	4.98	4.93	4.65	5.26
2	15	-	24.09	<b>15.46</b>	16.72	15.14	29.42	15.33	13.55	13.97	14.51
	25	-	34.27	25.45	27.01	<b>25.60</b>	37.80	23.42	21.74	20.77	24.40
	50	-	57.55	<b>50.40</b>	52.15	50.07	58.62	48.24	46.62	36.47	48.77
3	15	-	46.38	<b>34.23</b>	36.42	34.04	49.70	31.98	29.03	24.83	32.19
	25	-	67.58	<b>57.20</b>	59.13	56.94	68.72	52.91	50.93	39.34	55.89
	50	-	94.03	<b>91.31</b>	92.09	91.10	93.27	89.02	87.82	68.41	90.42
5	15	-	76.88	<b>65.17</b>	66.65	63.34	77.78	60.85	56.69	38.98	63.37
	25	-	94.59	<b>90.65</b>	91.22	89.81	94.18	87.65	85.47	60.91	89.96
	50	-	99.99	<b>99.95</b>	99.98	99.93	99.98	99.89	99.86	91.13	99.96
1	15	$BPG$			$BPG_u$			$K$		$K_u$	
		ASY	$BRT$	MC	ASY	$BRT$	MC	ASY	MC	ASY	MC
		4.52	5.47	5.44	11.06	12.75	5.30	4.41	5.47	10.37	5.31
	25	4.51	4.99	5.08	9.97	11.14	4.82	4.19	4.77	9.77	4.90
		4.48	4.83	4.91	8.02	8.37	4.63	4.38	4.48	7.82	4.94
		13.44	15.64	15.39	24.17	26.84	13.59	11.81	13.51	20.7	12.5
	50	23.95	25.45	25.07	33.86	35.66	21.13	20.98	22.60	30.55	19.07
		48.79	49.91	48.54	55.37	56.39	44.63	45.66	45.98	52.42	41.94
		28.37	32.16	31.34	41.13	44.47	26.92	22.78	25.49	34.47	21.96
	2	52.41	54.74	53.27	61.87	63.97	45.54	45.09	46.56	55.90	40.47
		89.53	90.07	88.91	91.11	91.65	85.59	86.55	85.67	89.08	81.66
		52.12	56.23	54.70	65.42	68.99	47.03	40.12	43.79	53.63	36.31
	3	84.81	86.17	84.58	89.37	90.48	77.60	75.14	76.00	82.62	67.70
		99.88	99.91	99.82	99.88	99.91	99.59	99.53	99.27	99.68	99.80
		99.88	99.91	99.82	99.88	99.91	99.59	99.53	99.27	99.68	99.80
1	15	Glejser			White			$C$		$H$	
		ASY <sub>F</sub>	ASY <sub>W</sub>	MC	ASY	MC		Tab.	MC	Tab.	MC
		5.42	6.58	5.47	4.97	5.1		12.52	5.37	12.97	5.04
	25	4.89	5.40	4.79	4.77	4.64		7.25	4.97	9.77	5.20
		4.61	4.89	4.84	5.41	5.26		4.34	4.92	3.41	5.26
		13.33	15.25	13.30	5.97	6.30		24.57	12.74	26.61	12.79
	50	21.36	23.36	21.17	5.82	5.47		24.07	18.38	32.86	20.63
		43.87	45.04	42.63	5.36	5.29		33.28	33.38	50.59	45.26
		26.59	29.86	26.11	6.65	6.89		37.88	22.28	46.20	27.63
	2	48.15	50.44	47.08	6.26	6.15		42.43	34.15	63.92	48.70
		85.09	85.86	83.92	5.31	5.27		64.77	65.01	90.58	87.30
		50.20	54.12	48.87	7.21	7.35		52.99	34.99	75.51	56.81
	3	81.06	82.68	79.44	6.79	6.52		63.55	54.60	93.13	85.91
		99.57	99.61	99.42	5.48	5.29		89.09	88.34	99.95	99.82
		99.57	99.61	99.42	5.48	5.29		89.09	88.34	99.95	99.82

technique to Hartley's criterion yields a very useful homoskedasticity test.

#### 5.4. Tests for break in variance

The model used is (2.1) with:  $T = 25, 50, 100$  and  $k = 6$ . The following specification for the error variance was considered:

$$\begin{aligned}\sigma_t^2 &= \sigma_1, & \text{if } t \leq \tau_0, \\ &= \sigma_1 + \delta, & \text{if } t > \tau_0,\end{aligned}$$

where  $\delta \geq 0$  and  $\tau_0$  is the break time (assumed unknown). The regressors and the regression coefficients parameters were chosen as in Section 5.3. Furthermore  $\alpha_0 = 1$ , and  $\delta$  and  $\tau_0$  were set so that:

$$\frac{(\sigma_1 + \delta)}{\sigma_1} = 1, 4, 16, \text{ and } \frac{\tau_0}{T} = .3, .5, .7.$$

We applied the MC versions of the standard tests  $GQ$ ,  $BPG$  and  $K$  (using artificial regressions on  $z_t = t$ ,  $1 \leq t \leq T$ ),  $W$ ,  $G$ ,  $RB$ ,  $S_F$ ,  $SKH$ ,  $S_N$  and  $HM$  tests, as well as the proposed combined tests  $F_{\times}(BPG; \hat{J}_{(4)})$ ,  $F_{\min}(BPG; J)$ ,  $F_{\times}(GQ; \hat{K}_{(4)})$ ,  $F_{\min}(GQ; K)$ . For each one of the combined tests, we considered two possible “windows” ( $J$ ,  $K$ ). The first one is a relatively uninformative “wide” window:

$$J^A = \{1, \dots, T-1\}, \quad (5.16)$$

$$K^A = S_1(T, T_2, k+1, T-T_2-k-1), \quad (5.17)$$

with  $T_2 = \lceil T/5 \rceil$ . The second set of windows were based on a predetermined interval around the true break-date, namely we considered:

$$J^S = \{L_0, L_0 + 1, \dots, U_0\}, \quad (5.18)$$

$$K^S = S_1(T, T_2, \tau_0^L(k), \tau_0^U(k)), \quad (5.19)$$

where  $T_2 = \lceil T/5 \rceil$ ,

$$\tau_0^L(k) = \max \{k+1, \tau_0 - I[T/5]\}, \quad \tau_0^U(k) = \min \{T-k-T_2, \tau_0 + I[T/5]\}. \quad (5.20)$$

This yields the statistics  $F_{\times}(BPG; \hat{J}_{(4)}^i)$ ,  $F_{\min}(BPG; J^i)$ ,  $F_{\times}(GQ; \hat{K}_{(4)}^i)$ ,  $F_{\min}(GQ; K^i)$ ,  $i = A, S$ . The results are reported in Table 10.

As expected, the MC versions of all the tests achieve perfect size control. The results on relative power across tests agree roughly with those from the other experiments. Two points are worth noting. First, a remarkable finding here is the good performance of the Szroeter-type MC tests, which outperform commonly used tests such as the  $BPG$  and the  $GQ$  tests. For  $\tau_0/T = 0.3$ , the Bartlett test performs quite well in this experiment; note however that the test is implemented with  $T_1 = T_2 = I[T/3]$ . Second, the combined criteria perform well, and in several cases [especially, with  $T = 50, 100$ ] exhibit the best performance. Among these tests, product-type combined criteria perform better than min-type. The combined GQ criteria clearly dominate the standard GQ; the

Table 10. Break in variance at unknown points

$T = 25$		$H_0$	$\tau_0/T = .3$			$\tau_0/T = .5$		$\tau_0/T = .7$	
$\sigma_2/\sigma_1$		1	2	4		2	4	2	4
$GQ$	$F$	3.9	8.8	16.1		21.5	73.1	17.1	62.5
$BPG$	ASY	4.1	9.8	17.1		22.9	52.9	33.0	77.0
	$BRT$	4.1	9.8	17.3		22.9	53.0	33.1	77.2
	MC	5.0	10.8	17.7		24.6	54.4	33.2	79.0
$K$	ASY	5.6	10.2	16.2		22.7	39.6	28.7	56.9
	MC	5.5	9.7	15.4		20.9	38.2	26.5	54.7
$W$	ASY	0.0	0.0	0.0		0.0	0.0	0.0	0.0
	MC	4.4	5.5	4.2		8.3	8.0	7.8	11.1
$G$	$ASY_F$	6.0	12.9	22.0		24.3	50.2	27.5	60.9
	$ASY_W$	7.7	15.1	25.8		27.9	55.0	31.6	65.5
	MC	5.9	12.1	20.2		22.2	47.9	26.1	58.4
$RB$	ASY	5.2	15.9	33.3		22.3	55.0	34.9	81.8
	MC	4.0	13.8	30.0		19.3	50.3	31.8	79.8
$S_F$	MC	7.0	<b>20.6</b>	<b>33.4</b>		<b>44.5</b>	83.1	39.0	83.8
$S_N$	ASY	4.9	17.7	29.0		37.8	68.4	45.1	87.2
	MC	5.9	19.2	31.1		38.9	69.4	46.4	86.6
$SKH$	Beta	5.8	20.6	33.5		44.3	76.6	49.8	90.0
	Bound	0.1	1.4	3.4		7.0	24.0	11.7	54.4
	Bound inconc.	39.1	75.1	82.6		81.7	74.8	76.7	45.1
	MC	6.2	19.9	30.9		42.6	73.5	47.7	<b>88.6</b>
$HM$	Bound	0.0	0.5	1.2		4.0	23.4	1.9	17.0
	Bound inconc.	22.8	51.6	61.4		78.4	75.1	71.2	79.3
	MC	6.1	20.2	29.8		42.6	<b>86.8</b>	36.9	82.7
Tests maximized over the whole sample									
$F_{\times}(BPG; \hat{J}_{(4)}^A)$	MC	5.2	11.1	14.0		24.1	46.2	40.9	80.6
$F_{\min}(BPG; \hat{J}_{(4)}^A)$	MC	4.4	8.9	10.7		18.4	32.2	32.6	70.6
$F_{\times}(GQ; \hat{J}_{(4)}^A)$	MC	5.6	18.6	31.3		35.4	83.2	31.8	78.4
$F_{\min}(GQ; \hat{J}_{(4)}^A)$	MC	4.8	16.5	27.9		29.1	77.6	26.8	73.5
Tests maximized over a sub-sample									
$F_{\times}(BPG; \hat{J}_{(4)}^S)$	MC	5.5	8.8	17.3		33.6	68.2	<b>48.7</b>	86.9
$F_{\min}(BPG; \hat{J}_{(4)}^S)$	MC	5.5	8.8	17.3		33.6	68.2	<b>48.7</b>	86.9
$F_{\times}(GQ; \hat{J}_{(4)}^S)$	MC	5.6	19.2	32.5		35.4	83.2	32.2	72.3
$F_{\min}(GQ; \hat{J}_{(4)}^S)$	MC	5.6	19.2	32.5		35.4	83.2	32.2	73.2

Table 10. Break in variance at unknown points (continued)

$T = 50$		$H_0$	$\tau_0/T = .3$		$\tau_0/T = .5$		$\tau_0/T = .7$	
$\sigma_2/\sigma_1$		1	2	4	2	4	2	4
$GQ$	$F$	5.5	.35	59.9	71.2	99.8	57.1	99.3
$BPG$	ASY	5.3	32.6	59.3	67.0	96.5	73.7	99.6
	$BRT$	5.5	33.7	60.6	68.5	96.7	75.0	99.7
	MC	5.8	33.7	59.1	67.2	96.3	73.2	99.6
$K$	ASY	5.2	29.5	50.3	58.6	86.4	62.5	96.1
	MC	5.5	29.8	47.7	57.2	85.6	61.8	95.7
$W$	ASY	1.9	2.0	1.5	1.0	0.4	1.4	1.5
	MC	3.6	3.5	3.5	1.9	1.0	3.2	3.2
$G$	$ASY_F$	5.3	37.0	69.6	63.0	95.1	58.8	97.5
	$ASY_W$	5.8	39.7	71.6	64.8	95.5	60.9	97.7
	MC	5.4	35.9	68.0	60.6	93.8	56.5	97.1
$RB$	ASY	5.2	53.1	89.1	55.7	96.8	72.7	1.0
	MC	4.9	50.2	87.2	55.0	96.6	69.8	99.7
$S_F$	MC	7.2	47.0	70.3	82.9	99.5	73.3	99.4
$S_N$	ASY	5.6	47.7	72.3	80.8	99.1	82.3	99.9
	MC	6.2	45.5	70.8	77.4	98.1	80.4	99.7
$SKH$	Beta	9.0	58.5	79.5	87.9	99.8	87.1	1.0
	Bound	1.3	18.8	41.3	53.9	94.3	59.4	99.3
	Bound inconc.	22.8	64.5	52.0	42.7	5.7	36.0	0.7
	MC	6.8	46.7	71.9	81.9	99.0	81.1	99.8
$HM$	Bound	0.6	15.7	28.5	68.1	99.3	39.4	97.4
	Bound inconc.	16.8	45.5	49.3	29.4	0.7	49.3	2.6
	MC	5.6	35.5	54.7	<b>89.6</b>	99.9	69.3	99.2
Tests maximized over the whole sample								
$F_{\times}(BPG; \hat{J}_{(4)}^S)$	MC	5.3	16.1	26.1	45.3	86.7	72.5	99.5
$F_{\min}(BPG; \hat{J}_{(4)}^S)$	MC	5.5	12.7	18.0	31.7	66.5	62.6	98.7
$F_{\times}(GQ; \hat{J}_{(4)}^S)$	MC	5.6	56.8	<b>98.6</b>	79.1	<b>100</b>	73.3	99.7
$F_{\min}(GQ; \hat{J}_{(4)}^S)$	MC	6.0	50.0	98.2	71.6	99.8	67.2	99.4
Tests maximized over a sub-sample								
$F_{\times}(BPG; \hat{J}_{(4)}^S)$	MC	6.0	38.0	79.8	76.9	99.4	<b>81.7</b>	<b>99.9</b>
$F_{\min}(BPG; \hat{J}_{(4)}^S)$	MC	5.8	37.1	77.8	75.7	99.2	79.4	99.9
$F_{\times}(GQ; \hat{J}_{(4)}^S)$	MC	5.4	<b>60.7</b>	98.0	80.0	99.9	74.9	99.7
$F_{\min}(GQ; \hat{J}_{(4)}^S)$	MC	6.2	53.1	98.3	78.7	<b>100</b>	75.0	99.5



Table 10. Break in variance at unknown points (continued)

$T = 100$		$H_0$	$\tau_0/T = .3$			$\tau_0/T = .5$		$\tau_0/T = .7$	
$\sigma_2/\sigma_1$		1	2	4		2	4	2	4
$GQ$	$F$	5.6	65.8	87.7		97.4	<b>100</b>	91.9	<b>100</b>
$BPG$	ASY	5.2	65.8	92.3		96.0	100	95.6	100
	$BRT$	5.7	67.1	92.7		96.4	100	95.7	100
	MC	5.9	65.2	92.7		95.8	<b>100</b>	95.3	<b>100</b>
$K$	ASY	5.3	59.0	82.4		91.9	99.6	92.1	99.9
	MC	5.5	58.5	82.9		90.8	99.2	92.4	99.9
$W$	ASY	4.5	2.7	2.0		4.0	2.8	5.1	5.3
	MC	4.3	3.1	2.5		3.9	2.7	6.0	5.3
$G$	$ASY_F$	5.4	69.5	97.8		93.8	100	88.9	100
	$ASY_W$	5.4	70.6	98.0		94.0	100	89.2	100
	MC	5.1	68.3	97.1		92.4	<b>100</b>	88.5	<b>100</b>
$RB$	ASY	5.4	86.0	98.7		90.6	100	96.3	100
	MC	4.9	86.0	98.4		89.9	<b>100</b>	95.4	<b>100</b>
$S_F$	MC	5.5	74.5	91.9		98.2	<b>100</b>	95.3	<b>100</b>
$S_N$	ASY	4.9	78.4	96.4		98.4	100	97.7	100
	MC	5.5	78.5	95.6		98.0	<b>100</b>	97.9	<b>100</b>
$SKH$	Beta	8.1	85.7	98.1		98.9	100	98.4	100
	Bound	1.5	57.1	87.4		96.0	100	94.2	100
	Bound inconc.	12.3	34.7	11.8		3.5	0.00	4.9	0.00
	MC	5.3	77.4	95.2		98.2	<b>100</b>	97.6	<b>100</b>
$HM$	Bound	1.8	41.3	67.1		98.7	100	85.7	100
	Bound inconc.	8.9	34.7	21.5		1.2	0.00	12.1	0.00
	MC	5.2	61.1	79.6		<b>99.7</b>	<b>100</b>	93.6	<b>100</b>
Tests maximized over the whole sample									
$F_{\times}(BPG; \hat{J}_{(4)}^A)$	MC	5.9	26.3	52.3		82.5	<b>100</b>	94.6	<b>100</b>
$F_{\min}(BPG; \hat{J}_{(4)}^A)$	MC	5.9	20.0	33.1		64.5	99.3	90.0	<b>100</b>
$F_{\times}(GQ; \hat{J}_{(4)}^A)$	MC	5.6	91.5	<b>100</b>		97.3	<b>100</b>	96.0	<b>100</b>
$F_{\min}(GQ; \hat{J}_{(4)}^A)$	MC	5.4	89.1	<b>100</b>		96.6	<b>100</b>	94.3	<b>100</b>
Tests maximized over a sub-sample									
$F_{\times}(BPG; \hat{J}_{(4)}^S)$	MC	5.5	83.4	<b>100</b>		98.1	<b>100</b>	<b>98.2</b>	<b>100</b>
$F_{\min}(BPG; \hat{J}_{(4)}^S)$	MC	5.2	82.6	<b>100</b>		97.8	<b>100</b>	<b>98.2</b>	<b>100</b>
$F_{\times}(GQ; \hat{J}_{(4)}^S)$	MC	5.7	<b>94.4</b>	<b>100</b>		98.2	<b>100</b>	96.8	<b>100</b>
$F_{\min}(GQ; \hat{J}_{(4)}^S)$	MC	5.1	92.7	<b>100</b>		98.0	<b>100</b>	96.8	<b>100</b>

same holds true for the BPG-based tests, if the search window is not uninformative. Power increases importantly, where we consider the sup-tests maximized over the shorter, more informative window. These results have much to recommend the intuitively appealing combined tests, in association with the MC test method, in order to deal with problems of unknown shift in variance.

## 6. Conclusion

In this paper we have described how finite sample homoskedasticity tests can be obtained for a regression model with a specified error distribution. The latter exploit the MC test procedure which yields simulation-based exact randomized  $p$ -values irrespective of the number of replications used. The tests considered include tests for GARCH-type heteroskedasticity and sup-type tests against breaks in variance at unknown points. On observing that all test criteria are pivotal, the problem of “robustness to estimation effects” emphasized in Godfrey (1996) becomes irrelevant from our viewpoint. It is important to note that the general approach used here to obtain exact tests is not limited to the particular case of normal errors. In particular, the method proposed allows one to consider non-normal – possibly heavy-tailed (e.g., Cauchy) – error distributions, for which standard asymptotic theory would not apply.

The results of our simulation experiments suggest that Hartley-type and Szroeter-type tests seem to be the best choice in terms of power. Such tests have not gained popularity given the non-standard null distribution problem which we have solved here. We have introduced various MC combined tests, based on the minimum (sup-type tests) or the product (Fisher’s combination method) of a set of  $p$ -values, and demonstrated their good performance. Although the particular test statistics considered here are designed against a two-regime variance, it would be straightforward to implement, with similar MC methods, statistics aimed at detecting a larger number of variance regimes. Finally, in the context of conditional heteroskedasticity, we have solved the unidentified nuisance parameter problem relating to ARCH-M testing.

## References

- ALI, M. M., AND C. GIACCOTTO (1984): "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model," *Journal of Econometrics*, 26, 355–373.
- ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.
- (1999): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," Discussion paper, Cowles Foundation for Research in Economics at Yale University.
- ANDREWS, D. W. K., AND W. PLOBERGER (1995): "Admissibility of the Likelihood Ratio Test When a Parameter is Present Only under the Alternative," *The Annals of Statistics*, 23, 1383–1414.
- BARNARD, G. A. (1963): "Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25, 294.
- BARTLETT, M. S. (1937): "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London A*, 160, 268–282.
- BEG, R., M. SILVAPULLE, AND P. SILVAPULLE (1998): "Testing for ARCH in ARCH-in-Mean Model," Discussion paper, School of Business, La Trobe University, Bundoora, Australia.
- BERA, A. K., AND S. RA (1995): "A Test for the Presence of Conditional Heteroscedasticity Within ARCH-M Framework," *Econometric Reviews*, 14, 473–485.
- BEWLEY, R., AND H. THEIL (1987): "Monte Carlo Testing for Heteroscedasticity in Equation Systems," *Advances in Econometrics*, 6, 1–15.
- BICKEL, P. J. (1978): "Using Residuals Robustly I: Tests for Heteroscedasticity, Non-Linearity," *The Annals of Statistics*, 6, 266–291.
- BINKLEY, J. K. (1992): "Finite Sample Behaviour of Tests for Grouped Heteroscedasticity," *The Review of Economics and Statistics*, 74, 563–568.
- BOLLERSLEV, T., R. F. ENGLE, AND D. B. NELSON (1994): "ARCH Models," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle, and D. L. McFadden, chap. 49, pp. 2959–3038. North-Holland, Amsterdam.
- BREUSCH, T. S., AND A. R. PAGAN (1979): "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287–1294.
- BUSE, A. (1984): "Tests for Additive Heteroscedasticity: Goldfeld and Quandt Revisited," *Empirical Economics*, 9, 199–216.
- CHU, C.-S. J., AND H. WHITE (1992): "A Direct Test for Changing Trend," *Journal of Business and Economic Statistics*, 10, 289–299.

- COCHRAN, W. G. (1941): "The Distribution of the Largest of a Set of Estimated Variances As a Fraction of Their Total," *Annals of Eugenics*, 11, 351–361.
- CRIBARI-NETO, F., AND S. L. P. FERRARI (1995): "Bartlett-Corrected Tests for Heteroskedastic Linear Models," *Economics Letters*, 48, 113–118.
- CRIBARI-NETO, F., AND S. G. ZARKOS (1998): "Bootstrap Methods for Heteroscedastic Regression Models: Evidence on Estimation and Testing," working paper, Department of Economics, Southern Illinois University and Foundation of Economic and Industrial Research, Greece..
- DAGENAIS, M. G., AND J.-M. DUFOUR (1991): "Invariance, Nonlinear Models and Asymptotic Tests," *Econometrica*, 59, 1601–1615.
- DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press, New York.
- DAVIES, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.
- DAVIES, R. B. (1987): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.
- DEMOS, A., AND E. SENTANA (1998): "Testing for GARCH Effects: A One-Sided Approach," *Journal of Econometrics*, 86, 97–127.
- DUFOUR, J.-M. (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica*, 57, 335–355.
- (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475–494.
- DUFOUR, J.-M., AND M. G. DAGENAIS (1992): "Nonlinear Models, Rescaling and Test Invariance," *Journal of Statistical Planning and Inference*, 32, 111–135.
- DUFOUR, J.-M., AND L. KHALAF (1998): "Simulation Based Finite and Large Sample Inference Methods in Multivariate Regressions and Seemingly Unrelated Regressions," Discussion paper, C.R.D.E., Université de Montréal, 36 pages.
- (1999): "Finite Sample Tests in Seemingly Unrelated Regressions," Discussion paper, C.R.D.E., Université de Montréal, and GREEN, Université Laval, 19 pages.
- (2001): "Monte Carlo Test Methods in Econometrics," in *Companion to Theoretical Econometrics*, ed. by B. Baltagi, Blackwell Companions to Contemporary Economics, chap. 23, pp. 494–519. Basil Blackwell, Oxford, U.K.
- DUFOUR, J.-M., AND J. F. KIVIAET (1996): "Exact Tests for Structural Change in First-Order Dynamic Models," *Journal of Econometrics*, 70, 39–68.

- (1998): “Exact Inference Methods for First-Order Autoregressive Distributed Lag Models,” *Econometrica*, 66, 79–104.
- DUFOUR, J.-M., AND O. TORRÈS (1998): “Union-Intersection and Sample-Split Methods in Econometrics with Applications to SURE and MA Models,” in *Handbook of Applied Economic Statistics*, ed. by D. E. A. Giles, and A. Ullah, pp. 465–505. Marcel Dekker, New York.
- (2000): “Markovian Processes, Two-Sided Autoregressions and Exact Inference for Stationary and Nonstationary Autoregressive Processes,” *Journal of Econometrics*, 99, 255–289.
- DURBIN, J., AND G. S. WATSON (1950): “Testing for Serial Correlation in Least Squares Regression I,” *Biometrika*, 37, 409–428.
- DWASS, M. (1957): “Modified Randomization Tests for Nonparametric Hypotheses,” *Annals of Mathematical Statistics*, 28, 181–187.
- ENGLE, R. F. (1982): “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987–1007.
- (ed.) (1995): *ARCH: Selected Readings*, Advanced Texts in Econometrics. Oxford University Press, Oxford, U.K.
- ENGLE, R. F., D. F. HENDRY, AND D. TRUMBLE (1985): “Small-Sample Properties of ARCH Estimators and Test,” *Canadian Journal of Economics*, 18, 66–93.
- EVANS, M. A. (1992): “Robustness of Size of Tests of Autocorrelation and Heteroscedasticity to Nonnormality,” *Journal of Econometrics*, 51, 7–24.
- EVANS, M. A., AND M. L. KING (1985a): “Critical Value Approximations for Tests of Linear Regression Disturbances,” *Australian Journal of Statistics*, pp. 68–83, 27.
- (1985b): “A Point Optimal Test for Heteroscedastic Disturbances,” *Journal of Econometrics*, 27, 163–178.
- FAREBROTHER, R. W. (1987): “The Statistical Foundation of a Class of Parametric Tests for Heteroscedasticity,” *Journal of Econometrics*, 36, 359–368.
- FISHER, R. A. (1932): *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- FOLKS, J. L. (1984): “Combination of Independent Tests,” in *Handbook of Statistics, Volume 4, Nonparametric Methods*, ed. by P. R. Krishnaiah, and P. K. Sen, pp. 113–121. North-Holland, Amsterdam.
- GIACCOTTO, C., AND S. C. SHARMA (1988): “Jackknife Tests for Heteroscedasticity in the General Linear Model,” *Australian Journal of Statistics*, 30, 200–216.
- GLEJSER, H. (1969): “A New Test for Heteroscedasticity,” *Journal of the American Statistical Association*, 64, 316–323.

- GODFREY, L. G. (1978): "Testing for Multiplicative Heteroscedasticity,," *Journal of Econometrics*, 8, 227–236.
- GODFREY, L. G. (1988): *Misspecification Tests in Econometrics : The Lagrange Multiplier Principle and Other Approaches*. Cambridge University Press, Cambridge, UK.
- GODFREY, L. G. (1996): "Some Results on the Glejser and Koenker Tests of Heteroscedasticity," *Journal of Econometrics*, 72, 275–299.
- GOLDFELD, S. M., AND R. QUANDT (1965): "Some Tests for Heteroscedasticity," *Journal of the American Statistical Association*, 60, 539–547.
- GOURIÉROUX, C. (1997): *ARCH Models and Financial Applications*, Springer Series in Statistics. Springer-Verlag, New York.
- GOURIÉROUX, C., AND A. MONFORT (1995): *Statistics and Econometric Models, Volumes One and Two*. Cambridge University Press, Cambridge, U.K.
- GRIFFITHS, W. E., AND K. SUREKHA (1986): "A Monte Carlo Evaluation of the Power of Some Tests for Heteroscedasticity," *Journal of Econometrics*, 31, 219–231.
- HANSEN, B. E. (1996): "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.
- (1997): "Sample Splitting and Threshold Estimation," Discussion paper, Department of Economics, Boston College, Boston (MA).
- HARRISON, M. J. (1980): "The Small Sample Performance of the Szroeter Bounds Test for Heteroscedasticity and a Simple Test for Use When Szroeter's Test Is Inconclusive," *Oxford Bulletin of Economics and Statistics*, 42, 235–250.
- (1981): "A Comparison of the Bounds, Beta-approximate, and Exact Variants of Two Tests for Heteroscedasticity Based on Ordinary Least Squares Residuals," *The Economic and Social Review*, 12, 235–252.
- (1982): "Table of Critical Values for a Beta Approximation to Szroeter's Statistic for Testing Heteroscedasticity," *Oxford Bulletin of Economics and Statistics*, 44, 159–166.
- HARRISON, M. J., AND B. P. MCCABE (1979): "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals," *Journal of the American Statistical Association*, 74, 494–499.
- HARTLEY, H. O. (1950): "The Maximum  $F$ -Ratio a Short-Cut Test for Heterogeneity of Variances," *Biometrika*, pp. 308–312, 37.
- HARVEY, A. C. (1976): "Estimating Regression Models with Multiplicative Heteroscedasticity,," *Econometrica*, 44, 461–465.

- HARVEY, A. C., AND G. D. A. PHILLIPS (1974): "A Comparison of the Power of Some Tests for Heteroscedasticity in the General Linear Model," *Journal of Econometrics*, 2, 307–316.
- HONDA, Y. (1988): "A Size Correction to the Lagrange Multiplier Test for Heteroscedasticity," *Journal of Econometrics*, 38, 375–386.
- HONG, Y., AND R. D. SHEHADEH (1999): "A New Test for ARCH Effects and its Finite-Sample Performance," *Journal of Business and Economic Statistics*, 17, 91–108.
- IMHOF, P. J. (1961): "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426.
- JÖCKEL, K.-H. (1986): "Finite Sample Properties and Asymptotic Efficiency of Monte Carlo Tests," *The Annals of Statistics*, 14, 336–347.
- JUDGE, G. G., W. E. GRIFFITHS, R. CARTER HILL, H. LÜTKEPOHL, AND T.-C. LEE (1985): *The Theory and Practice of Econometrics*. John Wiley & Sons, New York, second edn.
- KING, M. L. (1981): "A Note on Szroeter's Bound Test," *Oxford Bulletin of Economics and Statistics*, pp. 315–326, 43.
- KIVIET, J. F., AND J.-M. DUFOUR (1997): "Exact Tests in Single Equation Autoregressive Distributed Lag Models," *Journal of Econometrics*, 80, 325–353.
- KLÜPPELBERG, C., R. A. MALLER, M. VAN DE VYVER, AND D. WEE (2000): "Testing for Reduction to Random Walk in Autoregressive Conditional Heteroscedasticity Models," Discussion paper, Center for Mathematical Sciences, Munich University of Technology, and Department of Accounting and Finance, University of Western Australia, Nedlands, Australia.
- KOENKER, R. (1981): "A Note on Studentizing a Test for Heteroscedasticity," *Journal of Econometrics*, 17, 107–112.
- KOENKER, R., AND G. BASSETT (1982): "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43–61.
- LEE, J. H., AND M. L. KING (1993): "A Locally Most Mean Powerful Based Score Test for ARCH and GARCH Regression Disturbances," *Journal of Business and Economic Statistics*, 11, 17–27, Correction 12 (1994), 139.
- LEE, J. H. H. (1991): "A Lagrange Multiplier Test for GARCH Models," *Economics Letters*, 37, 265–271.
- LEROY, S. F. (1996): "Stock Price Volatility," in Maddala and Rao (1996), pp. 193–208.
- LUMSDAINE, R. (1995): "Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 13, 1–10.

- MACKINNON, J. G., AND H. WHITE (1985): "Some Heteroscedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305–325.
- MACNEILL, I. B. (1978): "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times," *The Annals of Statistics*, 6, 422–433.
- MADDALA, G. S., AND C. R. RAO (eds.) (1996): *Handbook of Statistics 14: Statistical Methods in Finance*. North-Holland, Amsterdam.
- MAEKAWA, K. (1988): "Comparing the Wald, LR and LM Tests for Heteroscedasticity in a Linear Regression Model," *Economic Letters*, 26, 37–41.
- MCCABE, B. P. (1986): "Testing for Heteroscedasticity Occurring At Unknown Points," *Statistics-Theory and Methods*, pp. 1597–1613, 15.
- MILLER, JR., R. G. (1981): *Simultaneous Statistical Inference (Second Edition)*. Springer-Verlag, New York.
- NEWHEY, W. K., AND J. L. POWELL (1987): "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847.
- PAGAN, A. R., AND A. D. HALL (1983): "Diagnostic Tests As Residual Analysis," *Econometric Reviews*, 2, 159–218, Comments and Reply, 219–254.
- PAGAN, A. R., AND Y. PAK (1993): "Testing for Heteroskedasticity," in *Handbook of Statistics 11: Econometrics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 489–518. North-Holland, Amsterdam.
- PALM, F. C. (1996): "GARCH Models of Volatility," in Maddala and Rao (1996), pp. 209–240.
- PEARSON, E. S., AND H. O. HARTLEY (1976): *Biometrika Tables for Statisticians*. Biometrika Trust, London, UK, 3rd edn.
- PEARSON, K. (1933): "On a Method of Determining Whether a Sample of Size  $n$  Supposed to Have Been Drawn from a Parent Population," *Biometrika*, 25, 379–410.
- RAMSEY, J. B. (1969): "Tests for Specification Errors in the Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society Series B*, 31, 350–351.
- RIVEST, L.-P. (1986): "Bartlett's, Cochran's, and Hartley's Tests on Variances Are Liberal When the Underlying Distribution Is Long-Tailed," *Journal of the American Statistical Association*, 81, 124–128.
- SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. D. Intrilligator, pp. 827–879. North-Holland, Amsterdam.



- SHABAN, S. A. (1980): "Change Point Problem and Two-Phase Regression: An Annotated Bibliography," *International Statistical Review*, 48, 83–93.
- SHARMA, S. C., AND C. GIACCOTTO (1991): "Power and Robustness of Jackknife and Likelihood-ratio Tests for Grouped Heteroscedasticity," *Journal of Econometrics*, 49, 343–372.
- SULLIVAN, M., AND D. E. A. GILES (1995): "The Robustness of ARCH/GARCH Tests to First-Order Autocorrelation," *Journal of Quantitative Economics*, 11, 35–61.
- SZROETER, J. (1978): "A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models," *Econometrica*, 46, 1311–1327.
- THEIL, H. (1971): *Principles of Econometrics*. John Wiley & Sons, New York.
- TIPPETT, L. H. (1931): *The Methods of Statistics*. Williams and Norgate, London.
- WESTFALL, P. H., AND S. S. YOUNG (1993): *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley & Sons, New York.
- WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.
- WILKINSON, B. (1951): "A Statistical Consideration in Psychological Research," *Psychology Bulletin*, 48, 156–158.
- ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.